

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 10:50:55 ; Search time 16521.2 Seconds  
(without alignments)  
11538.043 Million cell updates/sec

Title: US-10-022-710-1  
Perfect score: 4398  
Sequence: 1 atggtgcgtgcattcagaa.....atggagacttagacatgtaa 4398

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.ov.\*
- 5: gb.pat.\*
- 6: gb.pl.\*
- 7: gb.pr.\*
- 8: gb.ro.\*
- 9: gb.sts.\*
- 10: gb.sy.\*
- 11: gb.vi.\*
- 12: gb.fun.\*
- 13: gb.in.\*
- 14: gb.mu.\*
- 15: gb.or.\*
- 16: gb.ov.\*
- 17: gb.pat.\*
- 18: gb.pl.\*
- 19: gb.ro.\*
- 20: gb.sts.\*
- 21: gb.sy.\*
- 22: gb.vi.\*
- 23: gb.fun.\*
- 24: gb.in.\*
- 25: gb.mu.\*
- 26: gb.or.\*
- 27: gb.ov.\*
- 28: gb.pat.\*
- 29: gb.pl.\*
- 30: gb.ro.\*
- 31: gb.sts.\*
- 32: gb.sy.\*
- 33: gb.vi.\*
- 34: gb.fun.\*
- 35: gb.in.\*
- 36: gb.mu.\*
- 37: gb.or.\*
- 38: gb.ov.\*
- 39: gb.pat.\*
- 40: gb.pl.\*
- 41: gb.ro.\*
- 42: gb.sts.\*
- 43: gb.sy.\*
- 44: gb.vi.\*
- 45: gb.fun.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4380.4	99.6	5724	9	AB051466 Homo sapi
2	3321.2	75.5	6112	10	BC063250 Mus muscu
3	2200.2	50.0	2421	9	BC033125 Homo sapi
4	2200.2	50.0	2422	9	BC019344 Homo sapi
5	1232.8	28.0	5669	9	AB023177 Homo sapi
6	1232.8	28.0	6373	6	AX079870 Sequence
7	1229.6	28.0	6378	6	AR338841 Sequence
8	1129.6	25.7	5032	6	AX818221 Sequence
9	1109.8	25.2	5287	6	AX780247 Sequence
10	580.6	13.2	3053	6	AX747433 Sequence
11	580.6	13.2	3053	9	AK092252 Homo sapi
12	527	12.0	119205	9	AC011231 Homo sapi
13	527	12.0	167462	9	AC027600 Homo sapi
14	362.4	8.2	224789	2	AC111611 Rattus no
15	360.8	8.2	259236	2	AC123741 Mus muscu
16	270.4	6.1	169928	9	AC019067 Homo sapi
17	239.2	5.4	194472	9	AC012000 Homo sapi
18	198.4	4.5	225453	2	AC097957 Rattus no
19	196.4	4.5	89212	9	AC016679 Homo sapi
20	189.2	4.3	215703	2	AC121276 Mus muscu
21	175.8	4.0	68894	2	AC100201 Mus muscu
22	166.4	3.8	258545	2	AC110536 Mus muscu
23	165	3.8	259236	2	AC123741 Mus muscu
24	163.4	3.7	277376	2	AC128972 Rattus no
25	154	3.5	62063	2	AC136706 Mus muscu
26	154	3.5	183783	9	AC017082 Homo sapi
27	137.6	3.1	256545	2	AC110536 Mus muscu
28	133.8	3.0	399	11	G61408 SHGC-85947
29	132.2	3.0	198944	2	AC115912 Mus muscu
30	132	3.0	221262	2	AC122394 Mus muscu
31	128.8	2.9	872	11	BV023652 S212P6059
32	124.2	2.8	95921	9	AC004141 Homo sapi
33	124.2	2.8	168008	9	AC142334 Pan trogl
34	124	2.8	215510	2	AC128337 Rattus no
35	124	2.8	263244	2	AC097540 Rattus no
36	119.4	2.7	198944	2	AC115912 Mus muscu
37	117.6	2.7	110000	2	Continuation (4 of
38	117.6	2.7	110000	2	Continuation (5 of
39	116.8	2.7	256124	2	AC130447 Rattus no
40	108.8	2.5	191439	10	AL627302 Mouse DNA
41	108.8	2.5	199140	2	BX005251 Mus muscu
42	103.6	2.4	1704	9	BC016980 Homo sapi
43	102	2.3	300201	2	AC107086 Rattus no
44	100.2	2.3	143751	9	AC004160 Homo sapi
45	98.2	2.2	85693	2	AC124117 Mus muscu

ALIGNMENTS

RESULT 1  
AB051466  
LOCUS AB051466 5724 bp mRNA linear PRI 07-FEB-2001  
DEFINITION Homo sapiens mRNA for KIAA1679 protein, partial cds.  
ACCESSION AB051466  
VERSION AB051466.1 GI:12697902  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Nagase, I., Kitano, R., Hattori, A., Kondo, Y., Okumura, K. and Ohara, O.  
TITLE Prediction of the coding sequences of unidentified human genes.  
XIX. The complete sequences of 100 new cDNA clones from brain which

code for large proteins in vitro  
DNA Res. 7 (6), 347-355 (2000)  
21082932  
11214970

REFERENCE  
PUBMED  
2 (bases 1 to 5724)  
Chara, O., Nagase, T. and Kikuno, R.  
Direct Submission  
Submitted (22-NOV-2000) Osamu Chara, Kazusa DNA Research Institute,  
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba  
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp,  
URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3913,  
Fax: 81-438-52-3914)

## FEATURES

## source

1. 5724  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="fh22954"  
/notes="vector: pBluescriptII SK plus"

## gene

1. 5724  
/genes="KIAA1679"

## CDS

<1. 4612  
/genes="KIAA1679"  
/notes="Start codon is not identified."  
/codon\_start=2  
/product="KIAA1679 protein"  
/protein\_id="BA521770.1"  
/db\_xref="GI:12697903"

translations="GMSHLSNCGESNRPPKRSFRVCDWSDHDLFWVEVSDWHCVL  
VPVARGVKEPTAECTAQLGHRVVRCTQKLNRTVANEICEHFALQPTQACLI  
PCPRDCVPSFLPWSNCKGKLOHRTAVIAPLFGLOCLNLTESRACDAISL  
PLGEEVTFSLKVPNSKCLPHLKEINPSGRVLDPSNERTVFKHOSKARHHS  
KSWAIEIGYTRQVSCRSQGNAMLSLCLQDSFPLTVOSCIKMPKDCETSONSSSPC  
SKTCSGLPLPFRSRNRVKNVAIGGKCEPPELLEKEACIVEGELLQCPKRYSWETS  
EMKEQVSLLEQDHPHWITVPGGGIQTREVIQAQSPVAAAARAKREVRPEKA  
LCVGPAPLSQNLNIPSTDCIVSSAWGLCTHENCHBPGKKGKFPTRQVLMEST  
GPAGHCPHLVSEPCEDPMYRLASEGICFPDHCGLGHRLLKKAQCNDEGDEVS  
SLCPVPPPRKSCIEIPCRMDLSEWTEWSQSCSNKNSDKQTRSRITLALAGE  
GKPCPSQALQHLRCHNDSCMLHWETSPWGCSEDTLVTALNATIWNAGEATCGV  
GIOTRVCVKSUVGWVTKRCPDSTPEVTRPCFLPKCKDCIVTAFSEWTPCPRMCO  
AGNAVQSRRIIIOAANGGCEPDTLYEBRECEVDSLCVYRKPKOKWSPCLILV  
ESVWQGTSSGSEGLQTRAVESCISDNRSAEMMECLQCTNGMFLLVQECTVPRE  
DCTFTAWSKTPCTCEATKRRRLQTKSRKKEQCDSDLYPLVETELCPCEEFIS  
QPYGNWSDCILPEGRREGLRVQADSKCEGLFRAVACSKDKNRPVDFPSCSS  
GYIEKCVIPCPFDKLSWSGSSGSCSICIGVIRSKWLKEKPYNGRPPFKDLK  
NOAQVHEAPVCSYCNQYSNVVEHWSCKINNELSLRCGGGTQSKIRCVNTADGEG  
GAVDSNLQNDIEIPETQSCSLMCPNCEVMSWELMSKCPQSCDPHTMORRTHLLRP  
SLNSRTCAEDSQVQCLNENCFQNLNLTWSTQLSNAPQCGQVTRFLLSCVSD  
GKPVMDQCSQHNLEKPFQMSIPCLVECVNQLSGMTAWTSCQTCGHGGRNTRF  
IIMPTQGEGRPCPTLTQKRTPTVPCYSWLVGNWSACKLEGGDCGEGVQIRLSLQV  
HSGISHAAGRVEDALGEMPPQDSILKQLCSVPQDGLHTEWSEWSTCELTIDGR  
SFETVGRQSRRTFIIOSFENQDSCPOQVLETRPCTGGKCYHYTWKASLWNNERTVW  
CORSDGVNVTGCSQSPARPAIRQCIACRKPFSYCTQGGVCGCEKGYTEIMSNGL  
DYCMKVGSEDKADVNLKSNRPVNSKIHDIKXGSLQPLDPLDGRVKIWIYGVSGG  
AFLIMIFLIFTSYLVYLVKKPKPHQSPFPQKPLIYLDGDLDM"

## ORIGIN

Query Match 99.6%; Score 4380.4; DB 9; Length 5724;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 4397; Conservative 0; Mismatches 1; Indels 6; Gaps 1;

QY 1 ATGFGCGCTGATTCAGAACTGAAACCGAAGTGTGGTGTGAAATGAAATGGAACAC 60  
DB 209 ATGFGCGCTGATTCAGAACTGAAACCGAAGTGTGGTGTGAAATGGAACAC 268  
QY 61 TTTCGCCCTTACGCTCTACAGAACAGAGCTTGCCTCATCTTCTCCCGGATTTGTGA 120  
DB 269 TTTCGCCCTTACGCTCTACAGAACAGAGCTTGCCTCATCTTCTCCCGGATTTGTGA 328  
QY 121 GTATCTGAGTTCTTACCATGTTCCAACTGTAGCAAGGAGATGTGGAGAAATTCAGCAT 180  
DB 329 GTATCTGAGTTCTTACCATGTTCCAACTGTAGCAAGGAGATGTGGAGAAATTCAGCAT 388

QY 181 AGAAGTGGCGGCTATAGCTCCCTCTCTTTGGTGGTTTGCATGTCCAAATCTGACT 240  
DB 389 AGAAGTGGCGGCTATAGCTCCCTCTCTTTGGTGGTTTGCATGTCCAAATCTGACT 448  
QY 241 GAGTCAAGAGCCCTGTGATGCTCCCATTTCTCTCTCTCTTTGGGGAAGAGGAATATACATTT 300  
DB 449 GAGTCAAGAGCCCTGTGATGCTCCCATTTCTCTCTCTCTTTGGGGAAGAGGAATATACATTT 508  
QY 301 AGCCTTAAGTTGGACCACTGAGTAATGAGAGAGTCCCTCATCTTAAAGAAATTAATCCA 360  
DB 509 AGCCTTAAGTTGGACCACTGAGTAATGAGAGAGTCCCTCATCTTAAAGAAATTAATCCA 568  
QY 361 AGCGGAAGAACTGTTCTGAGTTTAACTCTGATTCAAATGAGCGAGTCACTTTAAACAT 420  
DB 569 AGCGGAAGAACTGTTCTGAGTTTAACTCTGATTCAAATGAGCGAGTCACTTTAAACAT 628  
QY 421 CAAAGTTTACAAAGCACATCATCTTGAAGTCTTGGGCAATAGAGATAGTTTAAAC 480  
DB 629 CAAAGTTTACAAAGCACATCATCTTGAAGTCTTGGGCAATAGAGATAGTTTAAAC 688  
QY 481 CGGAGAGTTTCTGTACAAGAGTGTGGAACAAATGCTATGTTAAGCCCTTGGCTTCAA 540  
DB 689 CGGAGAGTTTCTGTACAAGAGTGTGGAACAAATGCTATGTTAAGCCCTTGGCTTCAA 748  
QY 541 GATTCTCTCCCATTTGACTGTTTCACTCTGATCATGCCAAAGACATGTGAAACCTCCAG 600  
DB 749 GATTCTCTCCCATTTGACTGTTTCACTCTGATCATGCCAAAGACATGTGAAACCTCCAG 808  
QY 601 TGGTCTCTCTGGAGCCCTGCTCCNAGACATGCGGTTCAGGAGTCTCTTGCACAGATTT 660  
DB 809 TGGTCTCTCTGGAGCCCTGCTCCNAGACATGCGGTTCAGGAGTCTCTTGCACAGATTT 868  
QY 661 AGGAGCAGAGCCCGGAACGTTGAACACATGGCTATTGGAGTGTGAAGAGTGTCTCTGAA 720  
DB 869 AGGAGCAGAGCCCGGAACGTTGAACACATGGCTATTGGAGTGTGAAGAGTGTCTCTGAA 928  
QY 721 CTTCTTGAAGAGAGGCTGCAATTTGTTGAAGAGAGAACTTCTGACAGCAATGTCCAGGAT 780  
DB 929 CTTCTTGAAGAGAGGCTGCAATTTGTTGAAGAGAGAACTTCTGACAGCAATGTCCAGGAT 988  
QY 781 TCCTTGGAGAACTTCTGAATGGAAGAAATGCAAGTCTCTCTCTCTCTCGAGCAGCAGGAT 840  
DB 989 TCCTTGGAGAACTTCTGAATGGAAGAAATGCAAGTCTCTCTCTCTCTCGAGCAGCAGGAT 1048  
QY 841 CCCCACCTGGCATGTGACGGGACCCGCTGTGGCGGTGGATTCAGACCCGGAGGTGTAC 900  
DB 1049 CCCCACCTGGCATGTGACGGGACCCGCTGTGGCGGTGGATTCAGACCCGGAGGTGTAC 1108  
QY 901 TGTGCCAGAGCGGTACACAGCAGCTCCGCACTGAGGCGCAAGGAAGTCTCTAGACCTGTG 960  
DB 1109 TGTGCCAGAGCGGTACACAGCAGCTCCGCACTGAGGCGCAAGGAAGTCTCTAGACCTGTG 1168  
QY 961 GAAAGAGCAATATGTGTGGGACCCCGCCGCTTCCCTCTCAGCTCTGCAATATCCCTTGC 1020  
DB 1169 GAAAGAGCAATATGTGTGGGACCCCGCCGCTTCCCTCTCAGCTCTGCAATATCCCTTGC 1228  
QY 1021 TCTAGCGATGATAGTATCTTCTGCTGAGCTGGGCTGTGCTCATCCATGAAACCTGT 1080  
DB 1229 TCTAGCGATGATAGTATCTTCTGCTGAGCTGGGCTGTGCTCATCCATGAAACCTGT 1288  
QY 1081 CATGAACCTCAGGGGAAAAAGGATTTAGAACGAGCGCAGCCCATGTCTCTCATGGAATCT 1140  
DB 1289 CATGAACCTCAGGGGAAAAAGGATTTAGAACGAGCGCAGCCCATGTCTCTCATGGAATCT 1348  
QY 1141 ACAGGCGCTGAGGCGATGCGCTCATTTGTTGGAGTCTGTTCTTGTGAGGATCCAATG 1200  
DB 1349 ACAGGCGCTGAGGCGATGCGCTCATTTGTTGGAGTCTGTTCTTGTGAGGATCCAATG 1408  
QY 1201 TGCTACCGATGCTGCGCATCAGAGGAGTCTGTTTCCCTGATCATGGAATAATGTGGCCTG 1260  
DB 1409 TGCTACCGATGCTGCGCATCAGAGGAGTCTGTTTCCCTGATCATGGAATAATGTGGCCTG 1468  
QY 1261 GGACATCGTATTCTGAAGCCCGCTCTGCCAAGATGACCGCGGAGAAATGTATCAGGAGT 1320

Db 1469 GGACATCGTATCTGAAGCGCTGTCAGAGATGACCGCGAGAGATGATCAGGGAGT 1528  
Qy 1321 CTTTGGCCAGTCTCCCTCTCTGAGAGAGTCTTTGAAATTCCTCGCGAATGGAC 1380  
Db 1529 CTTTGGCCAGTCTCCCTCTCTGAGAGAGTCTTTGAAATTCCTCGCGAATGGAC 1588  
Qy 1381 TGTGTGTGAGAGAGTGAAGGAGTGTCTATCTCTTCCAGTCTCTTCAATRAAAG 1440  
Db 1589 TGTGTGTGAGAGAGTGAAGGAGTGTCTATCTCTTCCAGTCTCTTCAATRAAAG 1648  
Qy 1441 TCAGATGGGAAACAGACAGCTCAAGAACTATCTCTGCACTGGCTGGGAGGTGAAAG 1500  
Db 1649 TCAGATGGGAAACAGACAGCTCAAGAACTATCTCTGCACTGGCTGGGAGGTGAAAG 1708  
Qy 1501 CCATGTCCCTCTAGTCAAGGTCTCCAGAGATCTGTTTGTGTATGACCAATCTCTGTATG 1560  
Db 1709 CCATGTCCCTCTAGTCAAGGTCTCCAGAGATCTGTTTGTGTATGACCAATCTCTGTATG 1768  
Qy 1561 CAGCTTCACTGGGAGATCGCTTGGGCCCTTGTCTGAGGACACATTTGTAATCTGCC 1620  
Db 1769 CAGCTTCACTGGGAGATCGCTTGGGCCCTTGTCTGAGGACACATTTGTAATCTGCC 1828  
Qy 1621 CTTAATGCAACATTTGGCTGGAATGGAGAGCCACGTGTGTGTAGGCAATTCAGACTCGG 1680  
Db 1829 CTTAATGCAACATTTGGCTGGAATGGAGAGCCACGTGTGTGTAGGCAATTCAGACTCGG 1888  
Qy 1681 AGAGTCTTCTGTCAAGAGTCAAGTGGGACAGTAATGACCAAAAGATGTCAGATTCT 1740  
Db 1889 AGAGTCTTCTGTCAAGAGTCAAGTGGGACAGTAATGACCAAAAGATGTCAGATTCT 1948  
Qy 1741 ACTCGACCTGAACTGTGCGCCCTGTTTCTCCATGCCAAAGAAAGTGTATTTGTGACT 1800  
Db 1949 ACTCGACCTGAACTGTGCGCCCTGTTTCTCCATGCCAAAGAAAGTGTATTTGTGACT 2008  
Qy 1801 GCTTTCACTGAGTGACACCTGCCCAGAGATGTGCCAAGCAGAGAAATGCCAGTAAAA 1860  
Db 2009 GCTTTCACTGAGTGACACCTGCCCAGAGATGTGCCAAGCAGAGAAATGCCAGTAAAA 2068  
Qy 1861 CAGTCTCATACAGATCATCATCAAGACAGCCATGAGGAGCCAGGATGCCAGAT 1920  
Db 2069 CAGTCTCATACAGATCATCATCAAGACAGCCATGAGGAGCCAGGATGCCAGAT 2128  
Qy 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCCCTGTGTCTGTATATCGTGGAG 1980  
Db 2129 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCCCTGTGTCTGTATATCGTGGAG 2188  
Qy 1981 CCACAGAAATGAGGCCCTTGTGATGTGAGAGTCTGTGAGAGTCTGTGAGAGTCTGTGAGAG 2040  
Db 2189 CCACAGAAATGAGGCCCTTGTGATGTGAGAGTCTGTGAGAGTCTGTGAGAGTCTGTGAGAG 2248  
Qy 2041 AGCAGTGAAGCCTGTGGAAGGGTTTACAAAGAGAGTGTCTCATGATCTCTGATGAC 2100  
Db 2249 AGCAGTGAAGCCTGTGGAAGGGTTTACAAAGAGAGTGTCTCATGATCTCTGATGAC 2308  
Qy 2101 AACCGGTGACAGAAATGATGGAATGCTTCAAGCAGACAAACGGCATGCTCTCTTGTG 2160  
Db 2309 AACCGGTGACAGAAATGATGGAATGCTTCAAGCAGACAAACGGCATGCTCTCTTGTG 2368  
Qy 2161 CAAAGATGACAGTCCCATGTGAGAGAGTGTGACCTTCTGCTTGTGTCAGATTTACG 2220  
Db 2369 CAAAGATGACAGTCCCATGTGAGAGAGTGTGACCTTCTGCTTGTGTCAGATTTACG 2428  
Qy 2221 CCGTGTCCAGAACTGTGAAGCCACAAAGATGAGCGCGCAGCTCACAGGGAAGAGC 2280  
Db 2429 CCGTGTCCAGAACTGTGAAGCCACAAAGATGAGCGCGCAGCTCACAGGGAAGAGC 2488  
Qy 2281 AGAAAGAGAGAAATGCCAGGATTTGACCTTTACCTTCTAGTGTGAGACAGAACTATGT 2340  
Db 2489 AGAAAGAGAGAAATGCCAGGATTTGACCTTTACCTTCTAGTGTGAGACAGAACTATGT 2548  
Qy 2341 CTTGTGATGAATTTATATCCCACTTATGGAATCTGTGATGCTTCTCCAGAA 2400

2549 CTTTGTGATGAATTTATATCCCAACTTATGAACTGGTCAGATTGATTTCTCCAGAA 2608  
Qy 2401 GGCAGAGGAGAGCTCACCAGGAGTCCGGTCAAGCAGACAGCAAGAAATGTGGAGAA 2460  
Db 2609 GGCAGAGGAGAGCTCACCAGGAGTCCGGTCAAGCAGACAGCAAGAAATGTGGAGAA 2668  
Qy 2461 GGCCTGCGCTTTCGAGCAGTACCTGTTCTGATAAAAATGGAAGACCTGTTGACCCCTCC 2520  
Db 2669 GGCCTGCGCTTTCGAGCAGTACCTGTTCTGATAAAAATGGAAGACCTGTTGACCCCTCC 2728  
Qy 2521 TTCTGACAGCTCTGTTTACATTCAGAAAAATGTGTCTATCTCCCTGCCATTTGATTGC 2580  
Db 2729 TTCTGACAGCTCTGTTTACATTCAGAAAAATGTGTCTATCTCCCTGCCATTTGATTGC 2788  
Qy 2581 TTTTAAAGCGATTTGTCTAGTTGGGGTCTTTCAGTTTCTTCTGGAATTTGGAGTGAGA 2640  
Db 2789 TTTTAAAGCGATTTGTCTAGTTGGGGTCTTTCAGTTTCTTCTGGAATTTGGAGTGAGA 2848  
Qy 2641 ATTGATCCAAATGGCTAAAAAGAAACCTTACATTTGAGGAGACCATGTCCTCCAACTG 2700  
Db 2849 ATTGATCCAAATGGCTAAAAAGAAACCTTACATTTGAGGAGACCATGTCCTCCAACTG 2908  
Qy 2701 GATCTCAAGAA-----TGAAGTACATGAGGAGTCCCATGTTACAGTGTGCAATCAG 2754  
Db 2909 GATCTCAAGAAATCAGGCTCAGGATCATGAGGAGTCCCATGTTACAGTGTGCAATCAG 2968  
Qy 2755 TATCTCTGTTTGTAGAACACTGTTCTTTCATGCAAAATCAACAAATGAGCTGAGTCCCTG 2814  
Db 2969 TATCTCTGTTTGTAGAACACTGTTCTTTCATGCAAAATCAACAAATGAGCTGAGTCCCTG 3028  
Qy 2815 CGTGTGAGGAGGAGAACATCTAGGAAATCAGATGTGTGAATCTGCGGATGGTGA 2874  
Db 3029 CGTGTGAGGAGGAGAACATCTAGGAAATCAGATGTGTGAATCTGCGGATGGTGA 3088  
Qy 2875 GGTGAGAGCTGATAGCAACCTGTGCAACAGGATGAAATTCCTCCAGAAACCCAGTCC 2934  
Db 3089 GGTGAGAGCTGATAGCAACCTGTGCAACAGGATGAAATTCCTCCAGAAACCCAGTCC 3148  
Qy 2935 TGTCTCTTATGTGTCCTCAATCAGTGTGTCTGTGTGAGTGGGACTTTGGAGCAATGC 2994  
Db 3149 TGTCTCTTATGTGTCCTCAATCAGTGTGTCTGTGTGAGTGGGACTTTGGAGCAATGC 3208  
Qy 2995 CCACAGTCTGAGTCCCGCAGACAAATGAGAGAACTGCGCACCTGCTTAAGACCATCA 3054  
Db 3209 CCACAGTCTGAGTCCCGCAGACAAATGAGAGAACTGCGCACCTGCTTAAGACCATCA 3268  
Qy 3055 CTGAACTCAAGGACTTGTGCTGAAGACTCACAGGTGAGCTTGCCTCTGATGAAAT 3114  
Db 3269 CTGAACTCAAGGACTTGTGCTGAAGACTCACAGGTGAGCTTGCCTCTGATGAAAT 3328  
Qy 3115 TGCTTCCAGTTCAGTACAACTTAAAGAGTGGAGCACAATGCGAGTGTGAGTGAACGCA 3174  
Db 3329 TGCTTCCAGTTCAGTACAACTTAAAGAGTGGAGCACAATGCGAGTGTGAGTGAACGCA 3388  
Qy 3175 CCCTGTGCTCAAGGCTCAGGACCCCTGCTGAAGCTGTGTGCTGAGTGTGAGTGAACGCA 3234  
Db 3389 CCCTGTGCTCAAGGCTCAGGACCCCTGCTGAAGCTGTGTGCTGAGTGTGAGTGAACGCA 3448  
Qy 3235 GTCAGCATGCAACATGTGAGCAGCAATAATTTGGAAGCCCGCAGAGATGAGCATTCOC 3294  
Db 3449 GTCAGCATGCAACATGTGAGCAGCAATAATTTGGAAGCCCGCAGAGATGAGCATTCOC 3508  
Qy 3295 TGCTTGTGTAATGCTGTGCTCAACTGTCTCAGCTCTCAGGCTGGAGCGCTTGGACAGAGTGT 3354  
Db 3509 TGCTTGTGTAATGCTGTGCTCAACTGTCTCAGCTCTCAGGCTGGAGCGCTTGGACAGAGTGT 3568  
Qy 3355 TCACAGACCTGTGGCCATGAGGTTCGAATGAGCGGACTCGATTTATTCATTTATTCACCAAC 3414  
Db 3569 TCACAGACCTGTGGCCATGAGGTTCGAATGAGCGGACTCGATTTATTCATTTATTCACCAAC 3628  
Qy 3415 CAAGGAGAGAGCGGCGCATGCCCGCAGAGCTTACCCAGAGAGAAACCTGCCAGAGTACC 3474  
Db 3629 CAAGGAGAGAGCGGCGCATGCCCGCAGAGCTTACCCAGAGAGAAACCTGCCAGAGTACC 3688

ILT 2  
3250  
IS  
NITIC  
SSION  
SION

```

location/Qualifiers
1. .6112
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:66597 IMAGE:6410178"
/clisuse_type="Brain, enriched mouse brain 12.5dp"
/clone_lib="NIH EMAP_F00"
/lab_host="DH10B"
/notes="Vector: pYX-ASC"
1. .6112
/gene="D130067t03Rik"
/notes="synonym: B93C082A18"
/db_xref="LocustID:210417"
/db_xref="MGI:2443925"
162. .4985
/gene="D130067t03Rik"
/codon_start=1
/product="D130067t03Rik protein"

```



GenCore version 5.1.6  
Copyright (C) 1993 - 2004 CompuGen Ltd.

cleic - nucleic search, using sw model  
n: February 23, 2004, 10:50:55 ; Search time 16521.2 Seconds  
(without alignments)  
11538.043 Million cell updates/sec  
US-10-022-710-1  
ct score: 4398  
nce: 1 atgggtgcgtgcattcagaa.....atggagacttagacatgtaa 4398  
ng table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
hed: 3470272 segs, 21671516995 residues  
l number of hits satisfying chosen parameters: 6940544  
num DB seq length: 0  
num DB seq length: 2000000000  
-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

ase : GenEmbl.\*

1: gb\_ba.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_cm.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4380.4	99.6	5724	9	AB051466	AB051466 Homo sapi
2	3321.2	75.5	6112	10	BC063250	BC063250 Mus muscu
3	2200.2	50.0	2421	9	BC033125	BC033125 Homo sapi
4	2200.2	50.0	2422	9	BC019344	BC019344 Homo sapi
5	1232.8	28.0	5669	9	AB023177	AB023177 Homo sapi
6	1232.8	28.0	6373	6	AX079870	AX079870 Sequence
7	1229.6	28.0	6378	6	AX338841	AX338841 Sequence
8	1139.6	25.7	5032	6	AX818221	AX818221 Sequence
9	1109.8	25.2	5287	6	AX780247	AX780247 Sequence
10	580.6	13.2	3053	6	AX747433	AX747433 Sequence
11	580.6	13.2	3053	9	AK092252	AK092252 Homo sapi
12	527	12.0	119205	9	AC011231	AC011231 Homo sapi
13	527	12.0	167462	2	AC027600	AC027600 Homo sapi
14	362.4	8.2	224789	2	AC111611	AC111611 Rattus no
15	360.8	8.2	259236	2	AC123741	AC123741 Mus muscu
16	270.4	6.1	189928	9	AC019067	AC019067 Homo sapi
17	239.2	5.4	194472	9	AC012000	AC012000 Homo sapi
18	198.4	4.5	225453	2	AC097957	AC097957 Rattus no
19	196.4	4.5	89212	9	AC016679	AC016679 Homo sapi
20	189.2	4.3	215703	2	AC121276	AC121276 Mus muscu
21	175.8	4.0	68894	2	AC100201	AC100201 Mus muscu
22	166.4	3.8	256545	2	AC110536	AC110536 Mus muscu
23	165	3.8	259236	2	AC123741	AC123741 Mus muscu
24	163.4	3.7	277376	2	AC128972	AC128972 Rattus no
25	154	3.5	62063	2	AC136706	AC136706 Mus muscu
26	154	3.5	183783	9	AC017082	AC017082 Homo sapi
27	137.6	3.1	258545	2	AC110536	AC110536 Mus muscu
28	133.8	3.0	399	11	G61408	G61408 SHGC-85947
29	132.2	3.0	198944	2	AC115912	AC115912 Mus muscu
30	132	3.0	221262	2	AC122394	AC122394 Mus muscu
31	128.8	2.9	872	11	BV023652	BV023652 S212P6059
32	124.2	2.8	95921	9	AC004141	AC004141 Homo sapi
33	124.2	2.8	168008	9	AC142334	AC142334 Pan trogl
34	124	2.8	215510	2	AC128337	AC128337 Rattus no
35	124	2.8	263244	2	AC097540	AC097540 Rattus no
36	119.4	2.7	198944	2	AC115912	AC115912 Mus muscu
37	117.6	2.7	110000	2	AC096315	AC096315 Continuation (4 of
38	117.6	2.7	110000	2	AC096315	AC096315 Continuation (5 of
39	116.8	2.7	256124	2	AC130447	AC130447 Rattus no
40	108.8	2.5	191439	10	AL627302	AL627302 Mouse DNA
41	108.8	2.5	195140	2	EX005251	EX005251 Mus muscu
42	103.6	2.4	1704	9	BC016980	BC016980 Homo sapi
43	102	2.3	300201	2	AC107086	AC107086 Rattus no
44	100.2	2.3	143751	9	AC004160	AC004160 Homo sapi
45	98.2	2.2	85693	2	AC124117	AC124117 Mus muscu

ALIGNMENTS

RESULT 1  
AB051466 5724 bp mRNA linear PRI 07-FEB-2001  
LOCUS AB051466 Homo sapiens mRNA for KIAA1679 protein, partial cds.  
DEFINITION Homo sapiens mRNA for KIAA1679 protein, partial cds.  
ACCESSION AB051466  
VERSION AB051466.1 GI:12697902  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (sites)  
Nagase,T., Kikuno,R., Hattori,A., Kondo,Y., Okumura,K. and Ohara,O.  
Prediction of the coding sequences of unidentified human genes.  
XIX. The complete sequences of 100 new cDNA clones from brain which



1469	GGACATCGTATTCTGAAGCCGCTCTCCCAAGATGACCGGGAAGATATACGGAGT	1528	
1321	CTTTGCCAGTCCCTCTCTGAGAGAGAGTCTTTGAAATCCCTGCGGAATGGAC	1380	
1529	CTTTGCCAGTCCCTCTCTGAGAGAGTCTTTGAAATCCCTGCGGAATGGAC	1588	
1381	TGTGTCTGAGCGAGTGGACGGAGTGTCTATCCCTGTCAGTCTCTGTTCAATAA	1440	
1589	TGTGTCTGAGCGAGTGGACGGAGTGTCTATCCCTGTCAGTCTCTGTTCAATAA	1648	
1441	TCAGATGGGAAAAGACAGCCAGTCAAGAACTATCTGCACTGCTGGGAAGTGA	1500	
1649	TCAGATGGGAAAAGACAGCCAGTCAAGAACTATCTGCACTGCTGGGAAGTGA	1708	
1501	CCATGTCCCCCTAGTCAAGGCTCTCCAAAGACATGTTTGTATGACCAATCCTGT	1560	
1709	CCATGTCCCCCTAGTCAAGGCTCTCCAAAGACATGTTTGTATGACCAATCCTGT	1768	
1561	CAGCTTCACTGGGAGACATCGCTTGGGCGCTTGTCTGAGGACCATTTGTAATG	1620	
1769	CAGCTTCACTGGGAGACATCGCTTGGGCGCTTGTCTGAGGACCATTTGTAATG	1828	
1621	CTTAATGCAACCAATGCTGAATGGGAAGCCACGTGTGGTGTAGGCATTCAGT	1680	
1829	CTTAATGCAACCAATGCTGAATGGGAAGCCACGTGTGGTGTAGGCATTCAGT	1888	
1681	AGAGTCTTCTGTCAAGAGTCAAGTGGGCAAGTATGACCAAAAGATGTCAGAT	1740	
1889	AGAGTCTTCTGTCAAGAGTCAAGTGGGCAAGTATGACCAAAAGATGTCAGAT	1948	
1741	ACTCGACCTGAACTGTGGCGCCCTGTTTCTCCATGCAAAAAGAGCTGATTGT	1800	
1949	ACTCGACCTGAACTGTGGCGCCCTGTTTCTCCATGCAAAAAGAGCTGATTGT	2008	
1801	GCTTTCACTGAGTGAACCTGCGCCAGGATGTGCCAAGCAGGAATGCCACAGTAA	1860	
2009	GCTTTCACTGAGTGAACCTGCGCCAGGATGTGCCAAGCAGGAATGCCACAGTAA	2068	
1861	CAGTCTCGATACAGAACTCATCATCAAGAGCAGCCAAATGGAGGCCAGGAAT	1920	
2069	CAGTCTCGATACAGAACTCATCATCAAGAGCAGCCAAATGGAGGCCAGGAAT	2128	
1921	ACCTTATATGAGGAGAGAGTGTGAAGTCTTCTGTGCTCTGATATCGGTGGA	1980	
2129	ACCTTATATGAGGAGAGAGTGTGAAGTCTTCTGTGCTCTGATATCGGTGGA	2188	
1981	CCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAGTCTGTCTGGCAGGAATA	2040	
2189	CCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAGTCTGTCTGGCAGGAATA	2248	
2041	AGCAGTGAAGCTGTGGAAAGGGGTTAATAAAGAGTGTCTCATGATCTCTGAT	2100	
2249	AGCAGTGAAGCTGTGGAAAGGGGTTAATAAAGAGTGTCTCATGATCTCTGAT	2308	
2101	ACCCTGTCAAGCAATGATGGAATCCCTCAAGCAGCAAAAGGATGCTCTCTTGT	2160	
2309	ACCCTGTCAAGCAATGATGGAATCCCTCAAGCAGCAAAAGGATGCTCTCTTGT	2368	
2161	CAAGAATGCACTGCCATGTCAGAAAGTGTCACTTCACTGCTTGGTCCAGTTAG	2220	
2369	CAAGAATGCACTGCCATGTCAGAAAGTGTCACTTCACTGCTTGGTCCAGTTAG	2428	
2221	CCCTGTCTCCAGGACTGTGAAGCCACAAAGTAGCGGCGGACAGTCAAGGAAAG	2280	
2429	CCCTGTCTCCAGGACTGTGAAGCCACAAAGTAGCGGCGGACAGTCAAGGAAAG	2488	
2281	AGAAGAAGGAGAAATCCAGGATTTCTGAOCTTTACCTTATGTGGAGACAGAT	2340	
2489	AGAAGAAGGAGAAATCCAGGATTTCTGAOCTTTACCTTATGTGGAGACAGAT	2548	
2341	CCTTGTGATGATTTATATCCACCTTATGAACTGTCAGATTCATTTCTCCAGAA	2400	

  

2549	CCTTGTGATGATTTATATATCCCAACCTTATGGAACCTGCTCAGATTCATTCAGAA	2608	
2401	GGCAGAGGAGGAGCTTACCGAGGAGCTGCGGTACAGAGCAGCAAGAAATGTGAGAA	2460	
2609	GGCAGAGGAGGAGCTTACCGAGGAGCTGCGGTACAGAGCAGCAAGAAATGTGAGAA	2668	
2461	GGCCTGGGCTTTTGAGCAGTGTCTGTGTAATAAATGGAAGACCTGTTGACCCCTCC	2520	
2669	GGCCTGGGCTTTTGAGCAGTGTCTGTGTAATAAATGGAAGACCTGTTGACCCCTCC	2728	
2521	TTCTGAGCAGCTTCTGTTTACATCAAGAAATGTGTCATTCCTGCCCATTTGATTGC	2580	
2729	TTCTGAGCAGCTTCTGTTTACATCAAGAAATGTGTCATTCCTGCCCATTTGATTGC	2788	
2581	AAGTTAAGCGATTCGTTTGTGTTGGGCTTTGCAATTCATCTTGTGGAATTTGAGTGA	2640	
2789	AAGTTAAGCGATTCGTTTGTGTTGGGCTTTGCAATTCATCTTGTGGAATTTGAGTGA	2848	
2641	ATTGATCCAAATGGCTTAAAGAAACCTTACAAATGGAGGACCAACATGTCCTCAACTG	2700	
2849	ATTGATCCAAATGGCTTAAAGAAACCTTACAAATGGAGGACCAACATGTCCTCAACTG	2908	
2701	GATCTCAAGAA-----TACGTTACATGAGGAGCTCCCATGTTTACAGTGTGCAATCG	2754	
2909	GATCTCAAGAA-----TACGTTACATGAGGAGCTCCCATGTTTACAGTGTGCAATCG	2968	
2755	TATTCCTGGTGTGAGAACACTGTTCTTATGCAAAATCAACAAATGAGTGTGAGTCCCTG	2814	
2969	TATTCCTGGTGTGAGAACACTGTTCTTATGCAAAATCAACAAATGAGTGTGAGTCCCTG	3028	
2815	CGCTGTGAGGAGGAGAACCAATCTAGGAAATCAGATGTGAAATCTCGGATGGTGA	2874	
3029	CGCTGTGAGGAGGAGAACCAATCTAGGAAATCAGATGTGAAATCTCGGATGGTGA	3088	
2875	GGTGGAGCAGTGTGAGAACCTGTGCAACAGGATGAAATCCCGCAGAAACCCAGTCC	2934	
3089	GGTGGAGCAGTGTGAGAACCTGTGCAACAGGATGAAATCCCGCAGAAACCCAGTCC	3148	
2935	TGTTCTTCTTATGTTGCCCAATCAGTGTGTCATGTTCTGAGTGGGAGCTTTGGAGCAATGC	2994	
3149	TGTTCTTCTTATGTTGCCCAATCAGTGTGTCATGTTCTGAGTGGGAGCTTTGGAGCAATGC	3208	
2995	CCACATCTATCGGATCCCAACATGAGAGAGAACTGCGCACTGCTAGAGCAATCA	3054	
3209	CCACATCTATCGGATCCCAACATGAGAGAGAACTGCGCACTGCTAGAGCAATCA	3268	
3055	CTGAATCTCAGGAGCTTGTGTAAGACTCAGAGTGTGAGCTTGTCTCTGATGAAAT	3114	
3269	CTGAATCTCAGGAGCTTGTGTAAGACTCAGAGTGTGAGCTTGTCTCTGATGAAAT	3328	
3115	TGCTTCCAGTTCAGTATCAATCTAAAGAGTGTGAGCACTGCGAGTGTGAAACGCA	3174	
3329	TGCTTCCAGTTCAGTATCAATCTAAAGAGTGTGAGCACTGCGAGTGTGAAACGCA	3388	
3175	CCCTGTGGTCAAGGCTCAGGACCCCTGCTAAGCTGTGTGTCAGTGTGCAAGCA	3234	
3389	CCCTGTGGTCAAGGCTCAGGACCCCTGCTAAGCTGTGTGTCAGTGTGCAAGCA	3448	
3235	GTCAAGTGTGACCAATGTGAGCAGCAATAATTTGGAGAGCCCGCAGAGATGAGCATCC	3294	
3449	GTCAAGTGTGACCAATGTGAGCAGCAATAATTTGGAGAGCCCGCAGAGATGAGCATCC	3508	
3295	TGCTTGTGGATTCGCTGTTCACTGTGAGCTCTCAGGTTGGAGCGGCTTGGACAGAGTGT	3354	
3509	TGCTTGTGGATTCGCTGTTCACTGTGAGCTCTCAGGTTGGAGCGGCTTGGACAGAGTGT	3568	
3355	TCAAGAGCTGTGGCCCATGGAGTGTGAAATGAGCGGACTCGATTTATTCATTTGCCAACC	3414	
3569	TCAAGAGCTGTGGCCCATGGAGTGTGAAATGAGCGGACTCGATTTATTCATTTGCCAACC	3628	
3415	CAAGAGAGGAGCGGCAATGCCCAAGAGCTTACCCAGGAGAAACCTGCCAGTGCACC	3474	
3629	CAAGAGAGGAGCGGCAATGCCCAAGAGCTTACCCAGGAGAAACCTGCCAGTGCACC	3688	

```

3475 CCTCTACAGCTGGTCTCTGGCACTGGTCTGATGTAATTTGGAGGTGGAGACTGT 3534
|||||
3689 CCTCTCTACAGCTGGTCTCTGGCACTGGTCTGATGTAATTTGGAGGTGGAGACTGT 3748
|||||
3535 GGGGAAGAGTTCAGATCCGACGCTTTCTGATCGTGGTCCAGTGGTTCATATCTCAT 3594
|||||
3749 GGGGAAGAGTTCAGATCCGACGCTTTCTGATCGTGGTCCAGTGGTTCATATCTCAT 3808
|||||
3595 GCAGCTGGAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3654
|||||
3809 GCAGCTGGAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3868
|||||
3655 AAGCAGCTGTCTGTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 3714
|||||
3869 AAGCAGCTGTCTGTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCT 3928
|||||
3715 AGCAGATGTAATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATG 3774
|||||
3929 AGCAGATGTAATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATG 3988
|||||
3775 AGATCAAGACTTTTATATATATATATATATATATATATATATATATATATATAT 3834
|||||
3989 AGATCAAGACTTTTATATATATATATATATATATATATATATATATATATATAT 4048
|||||
3835 CTAGAACAACGCGCTTTGTACAGAGGCAATGTTATCACTACATGAGAAAGCAAGTCT 3894
|||||
4049 CTAGAACAACGCGCTTTGTACAGAGGCAATGTTATCACTACATGAGAAAGCAAGTCT 4108
|||||
3895 TGAACAATTAACGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3954
|||||
4109 TGAACAATTAACGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 4168
|||||
3955 GGCTGTCTCCCTCAGGCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4014
|||||
4169 GGCTGTCTCCCTCAGGCGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4228
|||||
4015 CCTTCTCTCTACTGTACACAGGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4074
|||||
4229 CCTTCTCTCTACTGTACACAGGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4288
|||||
4075 ATGAATCAAAATGGTTCCTGATTAATCTGATGAAAGTACCAAGGCTCAGAGGATAAAAA 4134
|||||
4289 ATGAGATCAAAATGGTTCCTGATTAATCTGATGAAAGTACCAAGGCTCAGAGGATAAAAA 4348
|||||
4135 GCTGATGTGAATCAAAATGGTTCCTGATTAATCTGATGAAAGTACCAAGGCTCAGAGGATA 4194
|||||
4349 GCTGATGTGAATCAAAATGGTTCCTGATTAATCTGATGAAAGTACCAAGGCTCAGAGGATA 4408
|||||
4195 TTTAAAGAGTGGTCTCTTCAACCACTTGATCCAGATGGCGGAGTAAATTTGGGTTTAT 4254
|||||
4409 TTTAAAGAGTGGTCTCTTCAACCACTTGATCCAGATGGCGGAGTAAATTTGGGTTTAT 4468
|||||
4255 GCGGTTTCAGGTGGCGCTTTTCTCATCATGATTTTCTTAATATTTTACTTCTACCTTGT 4314
|||||
4469 GCGGTTTCAGGTGGCGCTTTTCTCATCATGATTTTCTTAATATTTTACTTCTACCTTGT 4528
|||||
4315 TGAAGAAGCCAAACACATCAAGACACACCTCCCAAGAGGCTCTGACCTTAGCC 4374
|||||
4529 TGAAGAAGCCAAACACATCAAGACACACCTCCCAAGAGGCTCTGACCTTAGCC 4588
|||||
4375 TAGCATGGAGACTTAGCATGTAA 4398
|||||
4589 TAGCATGGAGACTTAGCATGTAA 4612
|||||

```

BC063250 6112 bp mRNA linear ROD 16-DEC-2003  
Mus musculus RIKEN cDNA D130067I03 gene, mRNA (cDNA clone MGC:66597  
IMAGE:6410178), complete cds.  
BC063250  
BC063250.1 GI:38649124

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT

MGC  
Mus musculus (house mouse)  
Mus musculus  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 6112)  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T.I., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schneitz, T.E., Brownstein, M.J., Udén, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Morley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Battifield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,  
Schnerich, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 6112)  
Strausberg, R.  
Direct Submission  
Submitted (02-DEC-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-3590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Preparation: M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [medpaxil.stanford.edu](mailto:medpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAX Plate: 125 Row: i Column: 17  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 27369657.  
Location/Qualifiers  
1. 6112  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="MGC:66597 IMAGE:6410178"  
/tissue\_type="Brain, enriched mouse brain 12.5dp"  
/clone\_lib="NIH BMAP\_F00"  
/lab\_host="DH10B"  
/lab\_note="Vector: pYX-ASC"  
1. 6112  
/gene="D130067I03Rik"  
/notes="synonym: B930082A18"  
/db\_xref="LocustID:210417"  
/db\_xref="MGI:2443925"  
162. 4985  
/gene="D130067I03Rik"  
/codon\_start=1  
/product="D130067I03Rik protein"

FEATURES  
source  
gene  
CDS

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 10:50:55 ; Search time 16521.2 Seconds  
(without alignments)  
11538.043 Million cell updates/sec

Title: US-10-022-710-1  
Perfect score: 4398  
Sequence: 1 atgggtcggtgcattcagaa.....atggagacttagacatgtaa 4398

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_rnd:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_hgtgo\_hum:\*  
40: em\_hgtgo\_mus:\*  
41: em\_hgtgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4380.4	99.6	5724	9	AB051466	AB051466 Homo sapi
2	3321.2	75.5	6112	10	BC063250	BC063250 Mus muscu
3	2200.2	50.0	2421	9	BC033125	BC033125 Homo sapi
4	2200.2	50.0	2421	9	BC019344	BC019344 Homo sapi
5	1232.8	28.0	5669	9	AB023177	AB023177 Homo sapi
6	1232.8	28.0	5669	9	AB023177	AB023177 Homo sapi
7	1229.6	28.0	6373	6	AX079870	AX079870 Sequence
8	1129.6	25.7	5032	6	AX818221	AX818221 Sequence
9	1109.8	25.2	5287	6	AX780247	AX780247 Sequence
10	580.6	13.2	3053	6	AX747433	AX747433 Sequence
11	527	12.0	119205	9	AC092252	AC092252 Homo sapi
12	527	12.0	119205	9	AC011231	AC011231 Homo sapi
13	362.4	8.2	224789	2	AC111611	AC111611 Rattus no
14	362.4	8.2	224789	2	AC111611	AC111611 Rattus no
15	360.8	8.2	259236	2	AC123741	AC123741 Mus muscu
16	270.4	6.1	169928	9	AC019067	AC019067 Homo sapi
17	239.2	5.4	194472	9	AC012000	AC012000 Homo sapi
18	198.4	4.5	225453	2	AC097957	AC097957 Rattus no
19	196.4	4.5	89212	9	AC016679	AC016679 Homo sapi
20	189.2	4.3	215703	2	AC121276	AC121276 Mus muscu
21	175.8	4.0	68894	2	AC100201	AC100201 Mus muscu
22	166.4	3.8	256545	2	AC110536	AC110536 Mus muscu
23	165	3.8	259236	2	AC123741	AC123741 Mus muscu
24	163.4	3.7	277376	2	AC128972	AC128972 Rattus no
25	154	3.5	62063	2	AC136706	AC136706 Mus muscu
26	154	3.5	183783	9	AC017082	AC017082 Homo sapi
27	137.6	3.1	256545	2	AC110536	AC110536 Mus muscu
28	133.8	3.0	399	11	G61408	G61408 SHGC-85947
29	132.2	3.0	198944	2	AC115912	AC115912 Mus muscu
30	132	3.0	221262	2	AC123394	AC123394 Mus muscu
31	128.8	2.9	872	11	BV023652	BV023652 S212P6059
32	124.2	2.8	95921	9	AC004141	AC004141 Homo sapi
33	124.2	2.8	168008	9	AC142334	AC142334 Pan trogl
34	124	2.8	18510	2	AC128337	AC128337 Rattus no
35	119.4	2.8	263244	2	AC097540	AC097540 Rattus no
36	117.6	2.7	198944	2	AC115912	AC115912 Mus muscu
37	117.6	2.7	110000	2	AC096315	Continuation (4 of
38	117.6	2.7	110000	2	AC096315	Continuation (5 of
39	116.8	2.7	256124	2	AC130447	AC130447 Rattus no
40	108.8	2.5	191439	10	AL627302	AL627302 Mouse DNA
41	108.8	2.5	199140	2	BX005251	BX005251 Mus muscu
42	103.6	2.4	1704	9	BC016980	BC016980 Homo sapi
43	102	2.3	300201	2	AC107086	AC107086 Rattus no
44	100.2	2.3	143751	9	AC004160	AC004160 Homo sapi
45	98.2	2.2	85693	2	AC124117	AC124117 Mus muscu

ALIGNMENTS

RESULT 1  
AB051466  
LOCUS AB051466 Homo sapiens mRNA for KIAA1679 protein, partial cds.  
DEFINITION Homo sapiens mRNA for KIAA1679 protein, partial cds.  
ACCESSION AB051466  
VERSION AB051466.1 GI:12697902  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Nagase,T., Kikuno,R., Hattori,A., Kondo,Y., Okumura,K. and Ohara,O.  
TITLE Prediction of the coding sequences of unidentified human genes.  
XIX. The complete sequences of 100 new cDNA clones from brain

code for large proteins in vitro  
DNA Res. 7 (6), 347-355 (2000)  
21082932  
11214970  
2 (bases 1 to 5724)  
Chara, O., Nagase, T. and Kikuno, R.  
Direct Submission  
Submitted (22-NOV-2000) Osemu Ohara, Kazusa DNA Research Institute,  
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba  
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp)  
URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3913,  
Fax: 81-438-52-3914

Location/Qualifiers  
1..5724  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="fh22954"  
/note="vector:pBluescriptII SK plus"  
1..5724  
/gene="XIAA1679"  
1..4612  
/gene="XIAA1679"  
/note="Start codon is not identified."  
/codon\_start=2  
/product="XIAA1679 protein"  
/protein\_id="BAB21770.1"  
/db\_xref="GI:12697903"  
/translation="GWTSLNCGESNRPKRSFRVCDWHSDLFPQWSDMHVCVL  
VPYARGVKPTAEVCTAQHGLQHRMVCIOKLARTVVAEIECHFALOPTBOACIL  
PCPRCVUSLFPNSCKSGKGLQHTRAVIAPLFGGLQCNLPRESBACDAPISC  
PLGESEYFSLKVGWPKSCRPLHLKEINPSRTVLDFNSDSNERVTPKHSYKAHHS  
KSWALEIGYQTRQVRSNDSQNALSLCLQDSFPLTVQSCIMPEDCTOWSWSPT  
SKTCSGSLLPGRVSNKXMAIGKGLQELKEKEACIVGELLQCPVRSYSTR  
EWKQCSLLLEQQPHVHTVPGVCGGIQTRVYCAQSPAAALRAKEVREVEKA  
LCVGPALPDLNLCSTDCITVSNWANGLCIENECHEPOCKKFRTRORHLEST  
GPAHCCHLVRSVPCEDPMCVRLASEGICFPDHCKGLGRHILKAVCONDRGDSG  
SLCPVPPPKRSKCIPECMOCVLEWTEMSSCSQSNKNSDKGKRSRTILALAGE  
GKPCPPSQAQALQHLNCDHSCMLQHWETISWPGSCSEDTLVNALNIGWNGEATCV  
GIQTRVFCVSHVGVQVMTKCPDSTRPETVRPFLPKCKDCIVTAFSEWTPCPMCO  
AGNATVQSRVRIIOEANGQCEPDLTYEERECEDVSLCPVYRFPKQWSPCLVY  
ESWQGITGSSBAGKLTQRAVSCISDNRSAEMECIKQTNMGPLLVBQCTVPCRE  
DCTFAWSTKFTPCSTNCATRSRLQTKRKEKCDSDLYPLVETELPCDEFTS  
OPVGNWSCILPEGRERPHGLRVQADSKEGEGLRFAVACSKNGRPVDPSPCSSS  
GIQKQKVIPEFDCKLSDWSWSGSCSSCGIGVIRSKWLKEXPYNGRCPKLDLK  
NOQVHEAVPYCSECNQYSWVHWSCKINNELSLRCGGGSRKIRCVNTADGG  
GAVDNLNQNDEIPEPTQSCSLMCPNECWSEWGLWPKQSCDPHTMQRTRHLRP  
SLNRTCAEDSVQPCILNENCFQYNLTEWSTQLESENAPCGQVTRLLSCVSD  
GKPVMDQCEQHNLEKQPMRMSIPCLVECVANQCSGMTAWECSQTCGHGRMSTRF  
IIMPTQEGRCRPTLTQKTCPTVPCYSWLGNWSACKLEGGDCGEVQIRSLCMV  
HSGSISHAAGRVEDALCGPQDSILKQLCSVPCPGDCHLTWSENSTCELTCDGR  
SFETVGRQSRSTFIIOQFENQDSCFQVLETRPCTGKCKYHWKASLWNNERTVW  
CORSQGVNVTGCSQAPAAIRQCIACRRKPFSTCTGGVCGEKGTELMRNGFL  
DYCMKPVGSEDKKADVNLGKNSGRPNVSKIHIDIFKWSLQPLDPDPRVKIWIYGVSGG  
AFILMIFLITFSLYLVCKPKPQSTPPOQKPLTLAYDGLDY"

FEATURES  
source

gene  
CDS  
181 AGAACTCGCGCGTGCATAGCTCCCTCTCTTTGGTGGTTTGCATGTCCTCAATCTGACT 240  
389 AGAACTCGCGCGTGCATAGCTCCCTCTCTTTGGTGGTTTGCATGTCCTCAATCTGACT 448  
241 GAGTCAAGAGCCCTGTGATGCTCCCATTTCTCTCTTGGGAAGAGGATATACATTT 300  
449 GAGTCAAGAGCCCTGTGATGCTCCCATTTCTCTCTTGGGAAGAGGATATACATTT 508  
301 AGCCTTAAGGTTGGACCATGGAGTAAATGCAGACTGCCTCATCTTTAAAGAAATTAATCCA 360  
509 AGCCTTAAGGTTGGACCATGGAGTAAATGCAGACTGCCTCATCTTTAAAGAAATTAATCCA 568  
361 AGCGGAAGAACTGTCTGGAATTTTAACTCTGATTTCAATAGCAGGATCACCTTTAAACAT 420  
569 AGCGGAAGAACTGTCTGGAATTTTAACTCTGATTTCAATAGCAGGATCACCTTTAAACAT 628  
421 CAAAGTTACAAAGCACATCATCATTCGAAGTCTTGGGCAATAGAGATAGGTTATCAAAACC 480  
629 CAAAGTTACAAAGCACATCATCATTCGAAGTCTTGGGCAATAGAGATAGGTTATCAAAACC 688  
481 CGGCAGGTTTCGTGTACAAAGAGTGAAGAACAAATGCTATGTTAAGCCTTTGCGCTTCAA 540  
689 CGGCAGGTTTCGTGTACAAAGAGTGAAGAACAAATGCTATGTTAAGCCTTTGCGCTTCAA 748  
541 GATTCTCTCCCATTTGACTGTTTCAAGTCTCTCATCATGCCCCAAAGACTGTGAAACCTCCACG 600  
749 GATTCTCTCCCATTTGACTGTTTCAAGTCTCTCATCATGCCCCAAAGACTGTGAAACCTCCACG 808  
601 TGTCTCTCTCGAGCCCTGCTCCAAAGACATCCGTTTCAAGGAGTCTCTTGCAGGATTT 660  
809 TGTCTCTCTCGAGCCCTGCTCCAAAGACATCCGTTTCAAGGAGTCTCTTGCAGGATTT 868  
661 AGGAGCAGAGCCGGAACCTGGAAGACATGGCTATTGAGAGTGGAAAGAGTGTCTGAA 720  
869 AGGAGCAGAGCCGGAACCTGGAAGACATGGCTATTGAGAGTGGAAAGAGTGTCTGAA 928  
721 CTTCCTTGAAGAAAGAGCCCTGCATTTGTTGAAGAGAACTTCTGCAGCAATGTCTCCAGGAT 780  
929 CTTCCTTGAAGAAAGAGCCCTGCATTTGTTGAAGAGAACTTCTGCAGCAATGTCTCCAGGAT 988  
781 TCTTGAGAACTTCTGAATGGAAAGAAATGCGAAGTCTCTCTCTCTCGAGCAGAGGAT 840  
989 TCTTGAGAACTTCTGAATGGAAAGAAATGCGAAGTCTCTCTCTCTCGAGCAGAGGAT 1048  
841 CCCACTGCGCATGTGACGGGACCCGCTGTGTGGCGTGGGATCCAGACCCGCGAGGTGTAC 900  
1049 CCCACTGCGCATGTGACGGGACCCGCTGTGTGGCGTGGGATCCAGACCCGCGAGGTGTAC 1108  
901 TGTGCCCAGAGCGTACAGCAGCTGCCGCACTGAGGGCCAAAGAACTCTCTAGACCTGTG 960  
1109 TGTGCCCAGAGCGTACAGCAGCTGCCGCACTGAGGGCCAAAGAACTCTCTAGACCTGTG 1168  
961 GAAAGGCATATGTGTGGGACCCGCTGTGTGGCGTGGGATCCAGCTCTGCAATATCCCTTGC 1020  
1169 GAAAGGCATATGTGTGGGACCCGCTGTGTGGCGTGGGATCCAGCTCTGCAATATCCCTTGC 1228  
1021 TCTACGAGCTGCATAGTATCTTCTCTGCTCAGCTCGGGGCTGTGCAATCCATGAAACTGT 1080  
1229 TCTACGAGCTGCATAGTATCTTCTCTGCTCAGCTCGGGGCTGTGCAATCCATGAAACTGT 1288  
1081 CATGAACCTCAGGGGAAAAAGGATTTAGAACAGGAGCGCCATCTCTCTCAATGGAATCT 1140  
1289 CATGAACCTCAGGGGAAAAAGGATTTAGAACAGGAGCGCCATCTCTCTCAATGGAATCT 1348  
1141 ACAGGCGCTCAGGGGATTCCTCTCTGCTGAGTGTCTCTCTGAGGATCCCAATG 1200  
1349 ACAGGCGCTCAGGGGATTCCTCTCTGCTGAGTGTCTCTCTGAGGATCCCAATG 1408  
1201 TGTACCGATGCTGAGGATCAGAAAGGATCTGTTTCCCTGATCATGGAATATGSCCTG 1260  
1409 TGTACCGATGCTGAGGATCAGAAAGGATCTGTTTCCCTGATCATGGAATATGSCCTG 1468  
1261 GGACATCGTATCTGAAGGCGGCTGCGCAAGATGACCGCGGAGAGATGTATCAGGAGT 1320

ORIGIN

Query Match 99.6%; Score 4380.4; DB 9; Length 5724;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 4397; Conservative 0; Mismatches 1; Indels 6; Gaps 1;  
1 ATGGTCGGTGCATTCAGAGCTGAAACCGAACTGTGGTTGCAATGAAATATGCAACAC 60  
209 ATGGTCGGTGCATTCAGAGCTGAAACCGAACTGTGGTTGCAATGAAATATGCAACAC 268  
61 TTGCGCTTCAGCTCCTCAGAACAGGCTTGCTCATTCCTTGTCCCGGATTTGTGA 120  
269 TTGCGCTTCAGCTCCTCAGAACAGGCTTGCTCATTCCTTGTCCCGGATTTGTGA 328  
121 GTATCTGAGTTCTTACATGATGTTCACTGTAGCAAGGATGTGGGAAGAAATTCAGCAT 180  
329 GTATCTGAGTTCTTACATGATGTTCACTGTAGCAAGGATGTGGGAAGAAATTCAGCAT 388

Db 1469 GGACATCGTATCTGAAGGCGCTGTGCCAGATGACCGCGAGAGATGTATCAGGAGT 1528  
QY 1321 CTTTGGCCAGTCCCTCTCTCTGTGAGAGAGTCTGTGAAATTCCTGCGCAATGGAC 1380  
Db 1529 CTTTGGCCAGTCCCTCTCTCTGTGAGAGAGTCTGTGAAATTCCTGCGCAATGGAC 1588  
QY 1381 TGTGTGCTGAGCGAGTGGACGGAGTGTCTATCCTGTTCCTGTTCCAGTCTCTGTTCAATAAAAAA 1440  
Db 1589 TGTGTGCTGAGCGAGTGGACGGAGTGTCTATCCTGTTCCTGTTCCAGTCTCTGTTCAATAAAAAA 1648  
QY 1441 TCAGATGGGAAAACAGACCGAGTCAAGAACTATCTGTGCACTGGCTGGGAGAGTGAAG 1500  
Db 1649 TCAGATGGGAAAACAGACCGAGTCAAGAACTATCTGTGCACTGGCTGGGAGAGTGAAG 1708  
QY 1501 CCATGTCCCTCTAGTCAAGGCTCTCCAGAGACATCGTTTGTGTAATGACCATTCCTGTATG 1560  
Db 1709 CCATGTCCCTCTAGTCAAGGCTCTCCAGAGACATCGTTTGTGTAATGACCATTCCTGTATG 1768  
QY 1561 CAGCTTCACTGGGAGACATCGCTTGGGGCCCTTGTCTGAGGACACATGGTAATGCTGCC 1620  
Db 1769 CAGCTTCACTGGGAGACATCGCTTGGGGCCCTTGTCTGAGGACACATGGTAATGCTGCC 1828  
QY 1621 CTTAATGCAACCAATTGGCTGAATGGAGAGACCAAGTGTGTGTAGGCAATTCAGACTCGG 1680  
Db 1829 CTTAATGCAACCAATTGGCTGAATGGAGAGACCAAGTGTGTGTAGGCAATTCAGACTCGG 1888  
QY 1681 AGAGTCTTCTGTGTCAGAGTCAAGTGGGCAAGTAATGACCAAAAGATGTCAGACTTCT 1740  
Db 1889 AGAGTCTTCTGTGTCAGAGTCAAGTGGGCAAGTAATGACCAAAAGATGTCAGACTTCT 1948  
QY 1741 ACTCGACTGAACTGTGGCCCTGTTTCTCCATGCAAAAGACTGTATGTGACT 1800  
Db 1949 ACTCGACTGAACTGTGGCCCTGTTTCTCCATGCAAAAGACTGTATGTGACT 2008  
QY 1801 GCTTTAGTGTAGTGACACCTGCCCAAGATGTGCCAAGCAGAAATGCCACAGTAAAA 1860  
Db 2009 GCTTTAGTGTAGTGACACCTGCCCAAGATGTGCCAAGCAGAAATGCCACAGTAAAA 2068  
QY 1861 CAGTCTCGATACAGAAATCATCCAGAGCAGCCAAATGGAGCCAGGAATGCCAGAT 1920  
Db 2069 CAGTCTCGATACAGAAATCATCCAGAGCAGCCAAATGGAGCCAGGAATGCCAGAT 2128  
QY 1921 ACCTTATATAGGAGAGAGAGTGTGAAGATGTTTCTGTGTCTGTATATCGTGTGAAG 1980  
Db 2129 ACCTTATATAGGAGAGAGAGTGTGAAGATGTTTCTGTGTCTGTATATCGTGTGAAG 2188  
QY 1981 CCACAGAAATGGAGCCCTTGATCTTATGTGCGCAGAGTCTGTGTGGCAGGGAATAACGGGC 2040  
Db 2189 CCACAGAAATGGAGCCCTTGATCTTATGTGCGCAGAGTCTGTGTGGCAGGGAATAACGGGC 2248  
QY 2041 AGCAGTGAAGCTGTGGAAGGGGTTACAAACAGAGCTGTCTCATGCATCTCTGATGAC 2100  
Db 2249 AGCAGTGAAGCTGTGGAAGGGGTTACAAACAGAGCTGTCTCATGCATCTCTGATGAC 2308  
QY 2101 AACCGGTGAGCAGAAATGATGGAATGCTTCAAGCAGACAAACGGCATGCTCTCTTTGTG 2160  
Db 2309 AACCGGTGAGCAGAAATGATGGAATGCTTCAAGCAGACAAACGGCATGCTCTCTTTGTG 2368  
QY 2161 CAAGATGCAAGTCCCATGTGCGAGAGACTGTACCTTCACTGCTTGGTCCAGTTTACG 2220  
Db 2369 CAAGATGCAAGTCCCATGTGCGAGAGACTGTACCTTCACTGCTTGGTCCAGTTTACG 2428  
QY 2221 CCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGTAGGGGGGACAGCTCAACAGGAAAAAGC 2280  
Db 2429 CCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGTAGGGGGGACAGCTCAACAGGAAAAAGC 2488  
QY 2281 AGAAGAGAGGAAATGCCAGATTCAGCTTTACCTTTAGTGGAGACAGAACTATGT 2340  
Db 2489 AGAAGAGAGGAAATGCCAGATTCAGCTTTACCTTTAGTGGAGACAGAACTATGT 2548  
QY 2341 CCTTGTGATGATTTATATCCCAACCTTATGAAACTGTCTCAGATTGCTTCTCCAGAA 2400

Db 2549 CCTTGTGATGAATTTATATCCCAACCTTATGAAAACCTGGTCAGATTGCATTTCTCCAGAA 2608  
QY 2401 GCGAAGGAGAGCTCACCGAGAGCTCGGGTACAAGCAGACAGACAAAGATGTGGAGAA 2460  
Db 2609 GCGAAGGAGAGCTCACCGAGAGCTCGGGTACAAGCAGACAGACAAAGATGTGGAGAA 2668  
QY 2461 GGCCTGCGCTTTGAGAGAGTGTCTGTATATAAATGGAAGACCTGTTGAGACCCCTCC 2520  
Db 2669 GGCCTGCGCTTTGAGAGAGTGTCTGTATATAAATGGAAGACCTGTTGAGACCCCTCC 2728  
QY 2521 TTCTGACAGCTCTGGTTACATTAAGAAAAATGTGTCAATTCCTGCTGCCATTTGATTGC 2580  
Db 2729 TTCTGACAGCTCTGGTTACATTAAGAAAAATGTGTCAATTCCTGCTGCCATTTGATTGC 2788  
QY 2581 AAGTTAAGCGATTTGTGTAGTGGGGTCTTGCAGTTTCACTTGTGGAATTTGGAGTGAGA 2640  
Db 2789 AAGTTAAGCGATTTGTGTAGTGGGGTCTTGCAGTTTCACTTGTGGAATTTGGAGTGAGA 2848  
QY 2641 ATTGCATCCAAATGCTAAAGAAACCTTACAAATGGAGGACGACCATGTCCCAAACTG 2700  
Db 2849 ATTGCATCCAAATGCTAAAGAAACCTTACAAATGGAGGACGACCATGTCCCAAACTG 2908  
QY 2701 GATCTCAAGAA-----TGAAGTACATGAGGAGTCCCATGTTTACAGTGAAGTCAATCAG 2754  
Db 2909 GATCTCAAGAAATCAGGCTCAGGTACATGAGGAGTCCCATGTTTACAGTGAAGTCAATCAG 2968  
QY 2755 TATTCCTGCTGTGTAGACACTGGTCTTCATGCAAAATCAAAATGAGCTGAGTCCCTG 2814  
Db 2969 TATTCCTGCTGTGTAGACACTGGTCTTCATGCAAAATCAAAATGAGCTGAGTCCCTG 3028  
QY 2815 CCGCTGAGAGAGGAAACAACTTAGGAAATCAGATGTGTAATCTGCGGATGGTGAA 2874  
Db 3029 CCGCTGAGAGAGGAAACAACTTAGGAAATCAGATGTGTAATCTGCGGATGGTGAA 3088  
QY 2875 GGTGAGCAGTGGATAGCAACCTGTGCAACAGAGTGAATTTCCCCAGAAACCCAGTCC 2934  
Db 3089 GGTGAGCAGTGGATAGCAACCTGTGCAACAGAGTGAATTTCCCCAGAAACCCAGTCC 3148  
QY 2935 TGTTCCTTATGTGTCCCAATGAGTGTGTCTGTGAGTGGGACTTTGGAGCAAAATGC 2994  
Db 3149 TGTTCCTTATGTGTCCCAATGAGTGTGTCTGTGAGTGGGACTTTGGAGCAAAATGC 3208  
QY 2995 CCACAGTCAATGCGATCCCCACACAAATCAGAGAGAACTCGGCCACTGCTAAGACATCA 3054  
Db 3209 CCACAGTCAATGCGATCCCCACACAAATCAGAGAGAACTCGGCCACTGCTAAGACATCA 3268  
QY 3055 CTGAATCAGAGACTGTGCTGAAGACTCAGAGTGCAGCTTGCCTCTGATGAAAT 3114  
Db 3269 CTGAATCAGAGACTGTGCTGAAGACTCAGAGTGCAGCTTGCCTCTGATGAAAT 3328  
QY 3115 TGTCTCCAGTTCAGTACAACTTAACAGAGTGGAGCACATGCGCAGCTGAGTGAACACGCA 3174  
Db 3329 TGTCTCCAGTTCAGTACAACTTAACAGAGTGGAGCACATGCGCAGCTGAGTGAACACGCA 3388  
QY 3175 CCGTGTGTTCAAGGCGTCAAGACCCGCTGTAACTCAGAGTGCAGCTTGCCTGAGTGAAGCA 3234  
Db 3389 CCGTGTGTTCAAGGCGTCAAGACCCGCTGTAACTCAGAGTGCAGCTTGCCTGAGTGAAGCA 3448  
QY 3235 GTCAGCATGCAAAATGTGAGCAGCATTAATTTGGAAGAGCCCAAGAGATGAGCATTTCC 3294  
Db 3449 GTCAGCATGCAAAATGTGAGCAGCATTAATTTGGAAGAGCCCAAGAGATGAGCATTTCC 3508  
QY 3295 TGTCTGTTGGAATGCGTGGTCAACTGTGAGCTCTCAGGGTGGACCGCTTGGACAGAGTGT 3354  
Db 3509 TGTCTGTTGGAATGCGTGGTCAACTGTGAGCTCTCAGGGTGGACCGCTTGGACAGAGTGT 3568  
QY 3355 TCACAGACCTGTGGCCATGGAGTGGATGAGCGGACTCGATTATCATTTATGCAAC 3414  
Db 3569 TCACAGACCTGTGGCCATGGAGTGGATGAGCGGACTCGATTATCATTTATGCAAC 3628  
QY 3415 CAAGGAGAGAGCGGCAATGCCCAAGAGCTTACCCAGGAGAAACCTGCCAGTACC 3474  
Db 3629 CAAGGAGAGAGCGGCAATGCCCAAGAGCTTACCCAGGAGAAACCTGCCAGTACC 3688





```
/protein_id="AAH63250.1"
/db_xref="GI:38649125"
/db_xref="LocusID:210417"
/db_xref="WGI:2443925"
/translation="MPLRSALVTHVSRMRKLFVLVLSLSOAAHLGRKDNQFLW
KTPGWRGACDCEPGGAGQSAVWCFHLEGHTSMNSNCDESSQPPKRSRCFVCDNHSO
LFQWESDTHKRLVLPDGAQGEPRPAVCEVTAQGLQHRTRVLAIFPLVYGGQCP
HFAQPTPEQAQCLIPCEPCDCEVSEFPCWSTCEGCKLQHRTRVLAIFPLVYGGQCP
NLTESACEAPVSEKPEYSEFSLKQVPMWCKRLPHLEKEDLSGRNIQDFSSDSNEQ
VILTHQYKAHHSQPDVDVIGFQTRQVMTCDGRNALLSLCVRDSFPLVQPCVMP
KDCETSPYSPWSPKSCIKRSSTLSPGVSRPNVNHIAIGGGQCEPELLEKETCAEG
EFLOPCRPYSHENKECOVSLILQHPDLWHETGPIGGGQIQTRBVVCAQSLNATI
ASIKYVSRVPEATLCLGPAPASQLCNVPCSMDCISVSNSTWPGVCFNCDPQKK
GFRMRQHLWESGMGRCPHLAESVPCDEPMCHRWLASBEGICIAHDKGCKGLHRL
KAVQNERGEVGGGLCPFPPEFMAECIEPCMDVSEWETWSSCSQSKNRSDG
KQTRSRILALAGEGGKCPSSQELQYRLCNDHSCTQLYWETSAAGSSENLTVLAD
NVTIGWGEATCGVIGTRKVFCTKSHVGWMTKRGCEPRTVPRPCFLPCKKDLV
TASEMTPCPCPGNTTQKSYRIIOEAANGQCEPDTLFEERECEDLSQPSY
RWKQWSSCILPESIRQRTGTSEAGKGLQTRAVSCISDDNQSAEMTECLQWNG
MPLVQECITPCADDCTFTPSRFTPCSKNEATQIRRLQTKGRKKECKQDASLYP
LVMEPCDCTFMSPHVGWNSACILPEGRKDAQGLWQGDDEKCEGVRFRATACS
NINGRPVDPCFNSGYIQACVIPCPCDKLSDWSWGSSSCGIGVRIIRKWLKE
KPYSGRCPKLDLKNQVHEAVPCISQDOYSWVHEWSPCKINNELSRPRCGRTQS
RRTCVSTADREGAVNRSLCNDQDAPOETQACSLCPSECVMSWETWSRCPQSDP
HAMQRRPHLLRPSLNSRTGDSQVPRCLLNENCFQFOYNLTWSTCOLSENVSGO
GUTRELLSCVRSKGSMDHCQRNLEKFPQMSIPECLVECVNQCISLWTTWEGSQ
TCQGGMRKIRFLIMPTEGGRQCEPTELQQRPCVPVPCISVWLNWNSACKLEGDC
EGVQVRSFCVHNGSISHTAVPEALCEVFPQEGHLQKLCVPCPDGCHITPWS
EMKSELCTCDGSRFTSGSRSTFIIQSPENQDSQVQVLETRPCCTGKCYHYIW
KASLMNNERTVMQSDGLNVTGGSPQARPAIRQCIPACKKPFYCTQGGVQCE
GVTETNRSRSGFLDYCMKVPGBSKKADVKNLSKRNRPVNSKIHDIFKMSQLPDLDP
GRYKMWYGVSGSFLIMIFVITSYLVCCKPKPQSTPRHQKPLILAYDGLDM"
705..842
/misc_feature
  /gene="D130067103Rik"
  /note="TSPI; Region: Thrombospondin type 1 repeats"
  /db_xref="CDD:smart00209"
  1179..1358
/misc_feature
  /gene="D130067103Rik"
  /note="TSPI; Region: Thrombospondin type 1 repeats"
  /db_xref="CDD:smart00209"
  3915..4067
/misc_feature
  /gene="D130067103Rik"
  /note="TSPI; Region: Thrombospondin type 1 repeats"
  /db_xref="CDD:smart00209"
  705..842
ORIGIN
Query Match      75.5%; Score 3321.2; DB 10; Length 6112;
Best Local Similarity 84.9%; Pred. No. 0;
Matches 3732; Conservative 0; Mismatches 663; Indels 3; Gaps 1;
QY      4 GTGGCTGCAATTCAGAGCTGAACCGCACTGTGGTTGCAATGAAATATCGCAACTTT 63
DB      588 GTGGCTGTCTTACAGAGCTCAATAGAACCAATGGTGTCCAATGAGATCTGTGAACACTTT 647
QY      64 GCCCTTCAGGCTCTCAAGAACAGCTTGCTCATTCCTTGTCCCGGGGATGTGTAGTA 123
DB      648 GCTCCTCAGCCCCCACAGAGCAAGCTGCTCATTCCTTGTCCCGGACTCGGTGTA 707
QY      124 TCTCAGTTCTTACCATGTCACACTGTAGCAAGGATGTCGGAAGAAATTCGACGATAGA 183
DB      708 TCTCAGTTCTCCCATGTCACCTGCGCCGAGGATGTGGGAAGAACTGCGACATAGA 767
QY      184 ACTCGCGGCTCATAGTCTCCCTCTCTTTGGTGGTTTGAATGTCCAAATCTGACTGAG 243
DB      768 ACTCGTGTGGCCATCGCGCCCTCTCTGTATGGAGGTCTGCAGTGTCCAAATCTCAGGAG 827
QY      244 TCAGAGCTGTGATCTCCCATCTCTCTCTCTTGGGAGAGGAAATATACATTAGC 303
DB      828 TCCAGAGCTCGCAGGCTCCAGTTCTGCTCTCTCTTGGGAAAGAGAAATATCTTTCAGC 887
QY      304 CTTAAGGTTGGACCATGAGTAAATGAGACTGTGCTCATCTTAAAGAAATTAATCAAGC 363
DB      888 CTAAGGTTGGACCATGAGCAATGTAGACTCTCTCACCTTAAGGAATCGACCTCAGC 947
```



Db 4188 GTCCCTGTGGAGAGGCCCTGTGTGGAGAGTACCCTTCCAGAGAAATCCTGGAAGCAG 4247

2Y 3661 CTGTGTTCTGTGCTTGTCCAGGAGACTGTCATTTAAACAGATGGTTCAGAGTGGAGCACA 3720

Db 4248 CTGTGCTCTGTGCTTGTCCAGGAGACTGCCACATACCCCATGGTCTGAGTGGACCAAG 4307

2Y 3721 TGTGAATTAACCTGCAATGATGAAGAGCTTTGAGACTGTGGCGGCCAGTCTAGATCA 3780

Db 4308 TGTGAGTAACCTGCAATGATGAAGAGCTTTGAGACTGTGGCGGCCAGTCTAGATCA 4367

2Y 3781 AGGACTTTTATAATTCAGTCTTTTGAAGAACCAAGACAGCTGCCCCCAACAGGTTCTAGAA 3840

Db 4368 AGGACTTCATATTCAGTCTTTTGAAGAACCAAGAGTGTGCCCCCAACAGGTTCTAGAA 4427

2Y 3841 ACAGCCCTTGTACAGAGGCAATGTTATCACTACATGAGAAAGCAAGTCTTGGAAAC 3900

Db 4428 ACAGGACATGTACAGGTTGCAATGTTATCACTACATGAGAAAGCAAGTCTTGGAAAC 4487

2Y 3901 AATAACGAACGAATGTATGTCAGCTGTCAGTGGCGTTTATGTCACAGGAGCTGC 3960

Db 4488 AATAATGAGCGGACTGTGTGTCAGGCTCAGATGGCCCTTAATGTACAGGTTGCTGC 4547

2Y 3961 TCCTCTCAGCCCGTCTGTGTCATTCGCGAGTGCATTCAGCCTGCGAAACCTTTC 4020

Db 4548 TCCTCTCAGCCCGTCTGTGTCATTCGCGAGTGCATTCAGCCTGCGAAACCTTTC 4607

2Y 4021 TCCTACTGTACAGAGGTGAGTGTGTGTGAGAGGGCTATACAGAGATATGAAA 4080

Db 4608 TCCTACTGTACAGAGGAGTGTGTGTGAGAGGGATACACAGAGATATGAAA 4667

2Y 4081 TCAATAGTTTCTCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4140

Db 4668 TCAATAGTTTCTCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4727

2Y 4141 GTGAAACCTTCTCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4200

Db 4728 GTGAAACCTTCTCGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4787

2Y 4201 GGATGGTCTTCTCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4260

Db 4788 GGATGGTCTTCTCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4847

2Y 4261 TCAGTGGGCTTTCTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320

Db 4848 TCAGTGGGCTTTCTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4907

2Y 4321 AAGCCAAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4380

Db 4908 AAGCCAAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4967

2Y 4381 GGAGACTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4398

Db 4968 GGAGACTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4985

RESULT 3

BC033125

LOCUS 2421 bp mRNA linear PRI 27-JUN-2002

DEFINITION Homo sapiens, clone IMAGE:3028838, mRNA, partial cds.

ACCESSION BC033125

VERSION BC033125.1 GI:21620030

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

1 (bases 1 to 2421)

Strausberg, R.

Submitted (25-JUN-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

## REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)

Steven Jones, Jennifer Asano, Ian Bosdet, Varon Butterfield,

Suanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,

Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo

Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven

Nees, Pawan Pandoh, Anna-Lilisa Prabhu, Parvaneh Saeedi, Jacqueline

Schein, Duane Smalios, Michael Smith, Lorraine Spence, Jeff Stott,

Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,

George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IBAL Plate: 43 Row: a Column: 6

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

## FEATURES

## source

Location/Qualifiers

1..2421

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="IMAGE:3028838"

/tissue type="Muscle; rhabdomyosarcoma"

/clone\_id="NTH MGC 17"

/lab\_hosts="DH10B-R"

/note="Vector: pOTB7"

<1..2224

/codon\_start=2

/product="Unknown (protein for IMAGE:3028838)"

/protein\_id="AAH33125.1"

/db\_xref="GI:21620031"

/translation="PHLKEINPSORTVLDPNSDSNERVTVKHYSKKAHHHSKSWALHII

## CDS

GTCTQCTSDGQNAWLSLCLQDPLTVQSCIMPDCDCTSOWMSQSCSKTCRSG

SLLPFRSRNRNVKHMAGGKECELEKEACIVEGELLQCCPRYSWRTSSEMKCVQ

SLLEQDPHWHTVPGVGGGIQTRVEYCAQSVFAAAALRAKEVREKALCVGPAP

LPOLCNIPGSTDVCIVSWAGWLCIHENCHDPQGGKGFRTORHLMESTGPGAGCP

HLVSPVEDPCMYRLASBGIQPDHKGKGLHRIILKAVCONDGRGDEVSGLCPVP

PPEKSKCEIPEMDCVLSEWTESSCSQSNKNSDGKQTRSTILLALAGEGKPCP

SQLQERHLNDHSCMQLHWETSPWGPCSEDTLVLTALNATIGNGENATCGVGQTRRV

FCVSRHQVQWTKKCPDSTRPVPCLPCKDCIVTAFSEWTPCPMCOAGNATVK

QSRVRIITQEAANGQCPDLYEERECEVDVSLCPVYRWKPKQSWPCILPESVWQVI

TGSSEAGKGLQTRAVSCI SDNRSAMECECLQTNMGMLLVQETVPCREDCTTAW

SKFTPCSTNCEATKSRRLQTKSRKKEKQDGLYPLVETELCPDCEFTSQPYGNWS

DCILPEGRREPHRLRVLQADSKEGCEGLRFRAVACSKNGRPVDPSPSCSSSKEMDGE

## ORIGIN

Query Match 50.0%; Score 2200.2; DB 9; Length 2421;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 336 GCCTCATCTTAAAGAAATTAATCCAGCGGAGAACTGTTCTGGATTTTAACCTGATTC 395

Db 1 GCCTCATCTTAAAGAAATTAATCCAGCGGAGAACTGTTCTGGATTTTAACCTGATTC 60

QY 396 AAATGAGCGAGTACCTTTAAACATCAAGTTACAAAGCACATCATCATTCGAAAGTCTTG 455

Db 61 AAATGAGCGAGTACCTTTAAACATCAAGTTACAAAGCACATCATCATTCGAAAGTCTTG 120

QY 456 GGCAATGAGATAGTGTATCAACCCGCGAGGTTTCGTGTACAAAGATGATGACAAAA 515

Db 121 GGCAATGAGATAGTGTATCAACCCGCGAGGTTTCGTGTACAAAGATGATGACAAAA 180

QY 516 TGCTATGTAAAGCCTTTGCCCTTCAAGATTCTCCCATTCAGTCTTCAGTCTCTCATCAT 575

Db 181 TGCTATGTAAGCCCTTTCCTTCAAGATTCCTCCCATGTAAGTCTGTCAGTCCCTCATCAT 240  
QY 576 GCCCAAGACGTGTGAACCTCCCAAGTGTCTCTCTGAGAGCCCTGCTCCCAAGACATGCCG 635  
Db 241 GCCCAAGACGTGTGAACCTCCCAAGTGTCTCTCTGAGAGCCCTGCTCCCAAGACATGCCG 300  
QY 636 TTCAGGAGTCTCTGACAGATTTAGAGCAGAGCCGGAACGTGAAGACATGGCTAT 695  
Db 301 TTCAGGAGTCTCTGACAGATTTAGAGCAGAGCCGGAACGTGAAGACATGGCTAT 360  
QY 696 TGGAGGTGAAGAGAGTGTCTGTAACCTTCTGAGAAAGAGCCCTGCATTTGTTGAAGGAGA 755  
Db 361 TGGAGGTGAAGAGAGTGTCTGTAACCTTCTGAGAAAGAGCCCTGCATTTGTTGAAGGAGA 420  
QY 756 ACTTCTGAGCAATGTCCTCCAGTATCTCTGAGAACTTCTGAATGGAAGAAATGCCAAGT 815  
Db 421 ACTTCTGAGCAATGTCCTCCAGTATCTCTGAGAACTTCTGAATGGAAGAAATGCCAAGT 480  
QY 816 CTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGACGGGACCCGCTGTGTGGCGG 875  
Db 481 CTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGACGGGACCCGCTGTGTGGCGG 540  
QY 876 TGGGATCCAGACCCGGGAGGTGTAATGTGCCAGAGCGTACACAGAGCTGCCGCACTGAG 935  
Db 541 TGGGATCCAGACCCGGGAGGTGTAATGTGCCAGAGCGTACACAGAGCTGCCGCACTGAG 600  
QY 936 GGCCAAGGAAGTCTCTAGACCTGTGGAAGAGCAATATGTGTGGGACCCGCTGTGTGGCGG 995  
Db 601 GGCCAAGGAAGTCTCTAGACCTGTGGAAGAGCAATATGTGTGGGACCCGCTGTGTGGCGG 560  
QY 996 CTCTCAGCTCTGCAATATCCCTTGCTTACGAGTGTGATATCTTCTGTCAGGCTG 1055  
Db 661 CTCTCAGCTCTGCAATATCCCTTGCTTACGAGTGTGATATCTTCTGTCAGGCTG 720  
QY 1056 GGGCTGTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1115  
Db 721 GGGCTGTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 780  
QY 1116 GCAGCGCATGTCTCATGGAATCTACAGGCGCTGCAGGCGATTCCTCATTTTGTGGA 1175  
Db 781 GCAGCGCATGTCTCATGGAATCTACAGGCGCTGCAGGCGATTCCTCATTTTGTGGA 840  
QY 1176 GTCTGTTCTGTGAGGATCCATGTTGCTACGATGGCTGCATCAGAGGATCTGTTT 1235  
Db 841 GTCTGTTCTGTGAGGATCCATGTTGCTACGATGGCTGCATCAGAGGATCTGTTT 900  
QY 1236 CCTCATCATGGAATGTGCGCTGGGACATGTAATCTGTAAGCGCTGCTGCCAGATGA 1295  
Db 901 CCTCATCATGGAATGTGCGCTGGGACATGTAATCTGTAAGCGCTGCTGCCAGATGA 960  
QY 1296 CCGCGGAGAGTGTATCAGGAGTCTTTGCCAGTTCCTCCAGTTCCTCTGAGAGGAATC 1355  
Db 961 CCGCGGAGAGTGTATCAGGAGTCTTTGCCAGTTCCTCCAGTTCCTCTGAGAGGAATC 1020  
QY 1356 TTGTGAATTCCTCCGAATGCAATGCTGTGTGCTGAGCGAGTGGACGGATGCTCATCTG 1415  
Db 1021 TTGTGAATTCCTCCGAATGCAATGCTGTGTGCTGAGCGAGTGGACGGATGCTCATCTG 1080  
QY 1416 TTCCAGTCTGTTCAAAATAAAATCAGATGGGAAACAGACCCAGGTCAAGAACTATCCT 1475  
Db 1081 TTCCAGTCTGTTCAAAATAAAATCAGATGGGAAACAGACCCAGGTCAAGAACTATCCT 1140  
QY 1476 GGCACTGGCTGGGAGGTGGAAAGCCATGTCCCTAGTCAGGCTCTCCAGAGCATCG 1535  
Db 1141 GGCACTGGCTGGGAGGTGGAAAGCCATGTCCCTAGTCAGGCTCTCCAGAGCATCG 1200  
QY 1536 TTGTGTAATGCAATTCCTGTATCAGCTTCACTGGGAGACATGCGCTTGGGCGCCCTTG 1595  
Db 1201 TTGTGTAATGCAATTCCTGTATCAGCTTCACTGGGAGACATGCGCTTGGGCGCCCTTG 1260  
QY 1596 TTCTGAGGACATTTGGTAACTGGCCCTTAATGCAACCATTTGGCTGGATGAGAGGCAAC 1655  
Db 1261 TTCTGAGGACATTTGGTAACTGGCCCTTAATGCAACCATTTGGCTGGATGAGAGGCAAC 1320

QY 1656 GTGTGTGTAGGCATTACAGACTCGGAGAGTCTTCTGTCTCAAGAGTCAAGTGGGCAAGT 1715  
Db 1321 GTGTGTGTAGGCATTACAGACTCGGAGAGTCTTCTGTCTCAAGAGTCAAGTGGGCAAGT 1380  
QY 1716 AATGACCAAAAGATGTCAGATTTACTCGACCTGAAACTGTGCGCCCTGTTTCTCCC 1775  
Db 1381 AATGACCAAAAGATGTCAGATTTACTCGACCTGAAACTGTGCGCCCTGTTTCTCCC 1440  
QY 1776 ATGCAAAAAGAGTGTATTTGTGACTGCTTTTCAGTGTAGTGGACACCTGCCAAGGATGTG 1835  
Db 1441 ATGCAAAAAGAGTGTATTTGTGACTGCTTTTCAGTGTAGTGGACACCTGCCAAGGATGTG 1500  
QY 1836 CCAAGCAGAAATGCCACAGTAAACAGTCTCTGATACAGATCATCATCAAGAGAGCAGC 1895  
Db 1501 CCAAGCAGAAATGCCACAGTAAACAGTCTCTGATACAGATCATCATCAAGAGAGCAGC 1560  
QY 1896 CAATGAGGCCAGGAATGCCAGATACCTTATATGAGGAGAGAGTGTGAAGATGTTTC 1955  
Db 1561 CAATGAGGCCAGGAATGCCAGATACCTTATATGAGGAGAGAGTGTGAAGATGTTTC 1620  
QY 1956 CTTGTGCTCTGTATATCGGTGAAAGCCACAGAAATGAGGCCCTTGATCTTAGTCCAGA 2015  
Db 1621 CTTGTGCTCTGTATATCGGTGAAAGCCACAGAAATGAGGCCCTTGATCTTAGTCCAGA 1680  
QY 2016 GTCTGTCTGCGAGGAATTAACCGGCGAGCAGTGAAGCTGTGGAAGGGTTCACAAACAG 2075  
Db 1681 GTCTGTCTGCGAGGAATTAACCGGCGAGCAGTGAAGCTGTGGAAGGGTTCACAAACAG 1740  
QY 2076 AGCTGTCTCATGCTCTCTGTATGACAAACCGGTGACAGAAATGATGGAATGCTCAAGCA 2135  
Db 1741 AGCTGTCTCATGCTCTCTGTATGACAAACCGGTGACAGAAATGATGGAATGCTCAAGCA 1800  
QY 2136 GACAAACCGCATGCTCTCTGTGCAAGAAATGCAAGTCCCATGTGCAAGAGATGTCAC 2195  
Db 1801 GACAAACCGCATGCTCTCTGTGCAAGAAATGCAAGTCCCATGTGCAAGAGATGTCAC 1860  
QY 2196 CTTCACTGCTTGGTCCAACTTTACCGCCCTGCTCCAGAACTGTGAAGCCACAAAAGTAG 2255  
Db 1861 CTTCACTGCTTGGTCCAACTTTACCGCCCTGCTCCAGAACTGTGAAGCCACAAAAGTAG 1920  
QY 2256 GCGGCGACAGCTCACAGGAAAGCAGAAAGAGAGAAATGCCAGGATCTGACCTTTA 2315  
Db 1921 GCGGCGACAGCTCACAGGAAAGCAGAAAGAGAGAAATGCCAGGATCTGACCTTTA 1980  
QY 2316 CCTCTAGTGGAGACAGAACTATGCTTGTGATGAATTTATATCCCACTTATGAGAA 2375  
Db 1981 CCTCTAGTGGAGACAGAACTATGCTTGTGATGAATTTATATCCCACTTATGAGAA 2040  
QY 2376 CTGGTCAGATTCATTTCTCCAGAGGAGGAGGCTCACCGAGACTGCGGGTACA 2435  
Db 2041 CTGGTCAGATTCATTTCTCCAGAGGAGGAGGCTCACCGAGACTGCGGGTACA 2100  
QY 2436 AGCAGACAGCAAGAAATGGAAGAGGCTGCGCTTTCAGCAGTAGCTGTTCTGTATAA 2495  
Db 2101 AGCAGACAGCAAGAAATGGAAGAGGCTGCGCTTTCAGCAGTAGCTGTTCTGTATAA 2160  
QY 2496 AAATGGAAGACCTGTGACCCCTCTTCTGACAGAGCTCTGTTA 2540  
Db 2161 AAATGGAAGACCTGTGACCCCTCTTCTGACAGAGCTCTGTTA 2205

## RESULT 4

BC019344  
LOCUS  
DEFINITION  
Homo sapiens KIAA1679 protein, mRNA (cdna clone IMAGE:3530946), partial cds.  
ACCESSION  
BC019344  
VERSION  
BC019344.1  
KEYWORDS  
GI:17939567  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 2422)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shermen, C.M., Schuler, G.D., Altshuler, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stappleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257  
 12477932  
 2 (bases 1 to 2422)  
 Strausberg, R.L.

Direct Submission  
 Submitted (13-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [gcgaps-remail.nih.gov](mailto:gcgaps-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada  
 info@bcsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Heliao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate; 15 Row; e Column: 12.

Location/Qualifiers  
 1. .2422  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3530946"  
 /tissue\_type="Muscle, rhabdomyosarcoma"  
 /clone\_lib="NIH\_MGC\_17"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 <1. .2422  
 /gene="KIAA1679"  
 /db\_xref="LocusID:80731"  
 <1. .2224  
 /gene="KIAA1679"  
 /codon\_start=2  
 /product="KIAA1679 protein"  
 /protein\_id="AAH19344.1"  
 /db\_xref="GI:17939568"  
 /db\_xref="LocusID:80731"

gene  
 CDS

/translation="PHLKEINPSGRVTLDPNSDSNERVTFKHOSYKAHHHSKWAIEI  
 GYTRVQSCSRSDGNAMSLCLQDFPLTVQSCIMPQKOCSTQMSWSPSCSTCSG  
 SLPLPGFSRNRNVKMAIGGKECIVVEGELIQCCPRYSWRTSEWKEQV  
 LLLEODPHWHTVPGVGGIOTREYVCAQSPVAAAALRAKEVSRPVEKALCVGPAP  
 LLSOLCNIPGCTDCIVSSNSAWGLCHENCHDPQKKKPRTRORHVLMESTGPAHCP  
 HLVEVPCEDPMCYRMASGICFPDQKCGELGHRLLKAVCONDREDVSGSLCPVPP  
 PPKAKSCEIFCRMDCVLSEWNSQSCSNKNSDGKQTRSTILLALAGEKPKFP  
 PQALQHLRNCNDSQMLHWETSPWGSCTDILVATNATIGNWGATCGVGIQTRRV  
 FVKVSHQVMTKRCPCDSTPFRFCFCKKDCIVTAFSEWTPCPRMCOAGNATVK  
 QSYRIIIOEAANGQECPTLYEERECEDVSLCPVYRMKPKQSCILVPEVWQGI  
 TGSEACGKGLQTRAVSISIDNKRASMECLKQTMGPLLVOECTVPCREDCTFTAW  
 SKPTPCSTNCEATKSRRLQTKSRKKEQDSLDVPLVETELPCDEPTSQPVGNWS  
 DCILPEGRPREHGLRVQADSKEGGLRPRVACSDKNRPNVDPSPSCSSSGKEMGE

ORIGIN

Query Match 50.0%; Score 2200.2; DB 9; Length 2422;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 336 GCCTCATCTTAAGAAATTAATCAACGCGAAGAACTGTTCTCGATTTTAACTCTGATTC 395  
 DB 1 GCCTCATCTTAAGAAATTAATCAACGCGAAGAACTGTTCTCGATTTTAACTCTGATTC 60  
 QY 396 AATGAGCGAGTCACTTTAAACATCAAGTTACAAAGCACATCATCATTCGAGTCTTG 455  
 DB 61 AATGAGCGAGTCACTTTAAACATCAAGTTACAAAGCACATCATTCGAGTCTTG 120  
 QY 456 GGCATAGAGATAGTTATCAACCCGCGAGTTCTGTGCAAGAGATGATGACAAA 515  
 DB 121 GGCATAGAGATAGTTATCAACCCGCGAGTTCTGTGCAAGAGATGATGACAAA 180  
 QY 516 TGCTATGTTAAGCCTTTGCTCAAGATTCCTCCCATTCAGTCTTCAGTCTTCGATCAT 575  
 DB 181 TGCTATGTTAAGCCTTTGCTCAAGATTCCTCCCATTCAGTCTTCAGTCTTCGATCAT 240  
 QY 576 GCCAAAGAGCTGTGAACCTCCAGTGTCTCTCTGAGCCCTCTGCTCCAAGACATGCCG 635  
 DB 241 GCCAAAGAGCTGTGAACCTCCAGTGTCTCTCTGAGCCCTCTGCTCCAAGACATGCCG 300  
 QY 636 TTCAGGAGTCTCTTGCCAGATTTAGGACGAGCGGAGCGGAGCGTGAAGACATGGCTAT 695  
 DB 301 TTCAGGAGTCTCTTGCCAGATTTAGGACGAGCGGAGCGGAGCGTGAAGACATGGCTAT 360  
 QY 696 TGGAGGTGGAAAGAGTGTCTGAACTTTTGAAGAAAGAGCGCTGCACTTTGTAAGAGAGA 755  
 DB 361 TGGAGGTGGAAAGAGTGTCTGAACTTTTGAAGAAAGAGCGCTGCACTTTGTAAGAGAGA 420  
 QY 756 ACTTCTGAGCAATGTCCAGGTATTCCTGAGAACTTCTGAAATGGAAGAATGCCAAGT 815  
 DB 421 ACTTCTGAGCAATGTCCAGGTATTCCTGAGAACTTCTGAAATGGAAGAATGCCAAGT 480  
 QY 816 CTCTCTCTCTGAGCAGCAGGATCCCACTGTCATGTGACGGACCCCTGTGTGGCGG 875  
 DB 481 CTCTCTCTCTGAGCAGCAGGATCCCACTGTCATGTGACGGACCCCTGTGTGGCGG 540  
 QY 876 TGGATCCAGACCCGGAGGTGTACTGTGCCAGAGCGTACCAAGACGTGCCGACTGAG 935  
 DB 541 TGGATCCAGACCCGGAGGTGTACTGTGCCAGAGCGTACCAAGACGTGCCGACTGAG 600  
 QY 936 GGCCAAAGAGTCTTAGACTGTGGAAGAGGCAATTATGTGGGACCCGCCCGTTGCC 995  
 DB 601 GGCCAAAGAGTCTTAGACTGTGGAAGAGGCAATTATGTGGGACCCGCCCGTTGCC 660  
 QY 996 CTCTCAGCTCTGCAATATCCCTTCTCTAGGACTGATAGTATCTTCTCGTGCAGCTG 1055  
 DB 661 CTCTCAGCTCTGCAATATCCCTTCTCTAGGACTGATAGTATCTTCTCGTGCAGCTG 720  
 QY 1056 GGGCTGTGCATCCATGAAACCTGTGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1115  
 DB 721 GGGCTGTGCATCCATGAAACCTGTGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 780  
 QY 1116 GCAGGCGCATGCTCTCATGGAATCTACAGGCGCTGCAGGCGCTTGCCTCATTTGGTGA 1175

Db 781 GCAGCGCCATGCTCTCATGGAATCTACAGGCGCTGCAGGCGATTCGCCCTCATTTGGTGA 840  
QY 1176 GTCTGTTCTTGTGAGGATCAATGTTGCTACCGATGCTGGCATCAAGAGGATCTGTTT 1235  
Db 841 GTCTGTTCTTGTGAGGATCAATGTTGCTACCGATGCTGGCATCAAGAGGATCTGTTT 900  
QY 1236 CCCTGATCATGGAATAATGTGGCTTGGACATCGTATTTCTGAAGCCCTCTGCCAGAAATGA 1295  
Db 901 CCCTGATCATGGAATAATGTGGCTTGGACATCGTATTTCTGAAGCCCTCTGCCAGAAATGA 960  
QY 1296 CCGGAGAGATGATFATCAGGAGTCTTTGCGCCAGTTCCCTCTCTCTCAGAGGAGTC 1355  
Db 961 CCGGAGAGATGATFATCAGGAGTCTTTGCGCCAGTTCCCTCTCTCTCAGAGGAGTC 1020  
QY 1356 TTGTGAATTCCTCCCGAATGGAATGTGTGCTGAGCGAGTGGACGGAGTGTCTATCTCTG 1415  
Db 1021 TTGTGAATTCCTCCCGAATGGAATGTGTGCTGAGCGAGTGGACGGAGTGTCTATCTCTG 1080  
QY 1416 TTCCAGTCTCTGTTCAATAAATACTCAGATGGGAACACAGACAGTCAAGACTATCTCT 1475  
Db 1081 TTCCAGTCTCTGTTCAATAAATACTCAGATGGGAACACAGACAGTCAAGACTATCTCT 1140  
QY 1476 GGCACTGGCTGGGGAAGTGAAGCCATGTCCTCCCTAGTCAGGCTCTCCAAAGCATCG 1535  
Db 1141 GGCACTGGCTGGGGAAGTGAAGCCATGTCCTCCCTAGTCAGGCTCTCCAAAGCATCG 1200  
QY 1536 TTTGTGTAATGACATTCCTGTATGACAGTTCATCTGGGAGACATCGCTTTGGGGCCCTTG 1595  
Db 1201 TTTGTGTAATGACATTCCTGTATGACAGTTCATCTGGGAGACATCGCTTTGGGGCCCTTG 1260  
QY 1596 TTCTGAGGACATCTGGTAACTGGCTTAAATGCAACCATTTGGCTGGAATGGAGAGCCAC 1655  
Db 1261 TTCTGAGGACATCTGGTAACTGGCTTAAATGCAACCATTTGGCTGGAATGGAGAGCCAC 1320  
QY 1656 GTGTGTTAGGATTCAGATCGGAGAGTCTTGTGTCAAGAGTCACTGGGACAAAGT 1715  
Db 1321 GTGTGTTAGGATTCAGATCGGAGAGTCTTGTGTCAAGAGTCACTGGGACAAAGT 1380  
QY 1716 AATGACCAAAAGATGTCCAGATTTCTCTGACCTGAAACTGTGGCCCTGTTTCTCCC 1775  
Db 1391 AATGACCAAAAGATGTCCAGATTTCTCTGACCTGAAACTGTGGCCCTGTTTCTCCC 1440  
QY 1776 ATGCAAAAAGACTGTATGTGACTGCTTTCAGTGAGTGACACCTCTGCCCAAGGATGTG 1835  
Db 1441 ATGCAAAAAGACTGTATGTGACTGCTTTCAGTGAGTGACACCTCTGCCCAAGGATGTG 1500  
QY 1836 CCAAGCAGGAATGCCAGATTAACAGTCTCGATACAGATCATCATCAAGAGAGCAGC 1895  
Db 1501 CCAAGCAGGAATGCCAGATTAACAGTCTCGATACAGATCATCATCAAGAGAGCAGC 1560  
QY 1896 CAATGGAGCCAGGAATGCCAGATACCTTATATGAGGAGAGAGTGTGAAGATGTTTC 1955  
Db 1561 CAATGGAGCCAGGAATGCCAGATACCTTATATGAGGAGAGAGTGTGAAGATGTTTC 1620  
QY 1956 CTTGTGTCCTGTATATCGGTGGAAGCCACAGAAATGAGAGCCCTTGCATCTTAGTGCAGA 2015  
Db 1621 CTTGTGTCCTGTATATCGGTGGAAGCCACAGAAATGAGAGCCCTTGCATCTTAGTGCAGA 1680  
QY 2016 GTCTGTCTGGCAGGAATAACGGCAGCAGTGAAGCTGTGGGAAGGGTTTACAACAAG 2075  
Db 1681 GTCTGTCTGGCAGGAATAACGGCAGCAGTGAAGCTGTGGGAAGGGTTTACAACAAG 1740  
QY 2076 AGCTGTCTCATGCATCTCTGTATGACAACCGGTGACGAGAAATGATGGAATGCTCTCAAGCA 2135  
Db 1741 AGCTGTCTCATGCATCTCTGTATGACAACCGGTGACGAGAAATGATGGAATGCTCTCAAGCA 1800  
QY 2136 GACAAACGGCATGCTCTCTTGTGCAAGATGACAGTCCCATGTGAGAGACTGCAC 2195  
Db 1801 GACAAACGGCATGCTCTCTTGTGCAAGATGACAGTCCCATGTGAGAGACTGCAC 1860  
QY 2196 CTTCACTCTTGGTCCAAAGTTTATGCGCCCTGCTCCACCAACTGTGGAAGCCACAAAAGTAG 2255

Db 1861 CTTCACTCTTGGTCCAAAGTTTACGCTGCTGCCAGAACTGTGAAGCCACAAAAGTAG 1920  
QY 2256 GGGGCGACAGCTCACAGGAAAAGCAGAAAAGAGAGAAATGCCAGGATCTTGACCTTTA 2315  
Db 1921 GGGGCGACAGCTCACAGGAAAAGCAGAAAAGAGAGAAATGCCAGGATCTTGACCTTTA 1980  
QY 2316 CCCTCTAGTGGAGACAGAACTATGCTCTTGTGATGAATTTATATCCCAACTTATGGAAA 2375  
Db 1981 CCCTCTAGTGGAGACAGAACTATGCTCTTGTGATGAATTTATATCCCAACTTATGGAAA 2040  
QY 2376 CTGGTCAGATGTCATTTCTCCAGAGGAGAGGAGGCTCACCGAGACTGCCGGTACA 2435  
Db 2041 CTGGTCAGATGTCATTTCTCCAGAGGAGAGGAGGCTCACCGAGACTGCCGGTACA 2100  
QY 2436 AGCAGACAGCAAAAGATGTGGAGAGGCTCGCTTTTCGACAGTAGCTGTTCTGTATAA 2495  
Db 2101 AGCAGACAGCAAAAGATGTGGAGAGGCTCGCTTTTCGACAGTAGCTGTTCTGTATAA 2160  
QY 2496 AAATGGAAGACTGTTGACCCCTCTCTGAGCAGCTCTGTTA 2540  
Db 2161 AAATGGAAGACTGTTGACCCCTCTCTGAGCAGCTCTGTTA 2205

RESULT 5  
LOCUS AB023177  
DEFINITION Homo sapiens mRNA for KIA0960 protein, partial cds.  
ACCESSION AB023177  
VERSION AB023177.2 GI:20521709  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE  
AUTHORS Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirose,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.  
TITLE Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro  
JOURNAL DNA Res. 6 (1), 63-70 (1999)  
MEDLINE 99246063  
PUBMED 10231032  
REFERENCE 2 (bases 1 to 5669)  
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.  
TITLE Direct Submission  
JOURNAL Submitted (04-FEB-1999) Otsu DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdhainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)  
COMMENT On May 9, 2002 this sequence version replaced gi:4589563.  
FEATURES  
Location/Qualifiers  
1..5669  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="h05779s1"  
/sex="male"  
/tissue\_type="brain"  
/clone\_lib="pBluescriptII SK plus"  
/dev\_stage="adult"  
/note="this sequence was obtained by subcloning of the DNA fragments derived from two cDNA clones (1 - 1401 was derived from th03753 and 1402 - 5669 was derived from h05779)."  
1..5669  
/gene="KIA0960"  
4..4510  
/gene="KIA0960"  
/codon\_start=2  
/product="KIA0960 protein"  
/protein\_id="BAA76804.2"  
/db\_xref="GI:20521710"



/translation="RBIACIQKDIDIPABDIIICEYFEPKPLLEQACLIIPCQDCDIIIVE  
 FSAWSECTSCGSLQHRVHVADPOFGSGCPNLTFBOVCQSPCEABELRYSLHV  
 GPMSTCSMPHSQVROARRGNKREKDRSGKVDPEARELLIKKRNRRNQNRQENK  
 YNDIQCYCTREVMCNKTKGADIISFCQOEXLPMTFQSCVITKEQVSESEWSPCS  
 KICHVDSPAGTRVTRIRQPISEKKEPEFEKPEKLSQGDGVVPCATYGNWRITE  
 WTECRVPLLSQDQKQEDTALCGGIGITREYICVQANENLLSGLTHKNKEAKPM  
 DLKLCCTGPTNTTQCHIEPTECEVPSWAMPCTYENCNDQOQKGFKLKRLRITN  
 EPTGSGVTGNCDFPHLEAIIPCEPACYDKWAVLGDCEPDNGKEGCGPTQVQEVVCI  
 SDEEYDRLCDRAIPPIVACADPCDKVLSIWTSSWSSHTCSGKTEGKQIRAR  
 SILAYAGEBGLIRCPNSSALQVRSNCHPCTVHMOCTGPMWQCTEDTSVSEFNIT  
 NGEASCSVMQTRKVIICVRNVGQVKKKPESLRPETVRPCLLPCKKDCIIVTPEYSD  
 WNSPSCSKEGSSIKQSRHVIIGLPANGRDCTDPLIEKACAEQAQCSYHWKT  
 HKWRRCOLPWSVQDQSPGAGQROARAITCKQDGGQAGHECLQVAGYPAL  
 TQACQIPCDQDQOLTSWSPSCNGCPVAVTRKTLVGSKKKCKKNCHLYPLIET  
 QVPCPKYNAQPVGNWSDCILPEKGVLLGMKVQDQIXECCQGYRYQAMACDNGR  
 LVTESRNSHGYIEEACIIPCSDDKLSWNSWRSKSGSVKVRKWLREKPYNG  
 GREPKLDHNOAQVVEVVPVCHSDNOYLWTEPMSICKVTFVNNRENCEGVQTRKY  
 RCQONTADGPESEHVEDYLDPEMPLGSRVCKLPQEDCVISWGPWQCVLPQWSS  
 PRORSADPTROPADEGRSPENAVEKEPNLNKNCHYDINVDWSTCQLSEKAVCGNG  
 IKTRMLDCVRSKGSVDLKYCEALGUEKNWQNTSCWVECPVNCQSDMSPSECSQT  
 CGLTGMIRRTVTQFGDGRPCPSLMDQSKPCPVKCYRWQYQWSPQVQEAQCYL  
 EGRTRRNIICSVSDGADDFSKVYDEFCADIELIIDKNMVLSESCSPQCPGDQYL  
 KXMSWSLQCLTCVNGEDLFGGIIQVRSRPVIOELNOHLCPBOMLETKSYCDQCY  
 EYKMASANKGSRVTWCORSQDINVTGCLVMSQPDADRSNPPCSPHPSYCSRTKI  
 CHCEEGYEVMSNSNLEACTLIPVVLPTMEDKRGDVKTSRAVHPTQPSNPAGRG  
 TWLQFPFGPGRUKTWYVGAAGFVLLIFIVSMIYLACKPKPKPQRRNNRLKPLTL  
 AYDGDADM"

ORIGIN

Query Match 28.0%; Score 1232.8; DB 9; Length 5669;  
 Best Local Similarity 57.7%; Pred. No. 0;  
 Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;  
 2Y 10 TGCATTGAGGTGAACCGAACTGTGGTGTGCAAAATGAAATATGCAACACTTTGCCCTT 69  
 2b 14 TGCATCAGAAAGACAAAGACATCTCTCGCGAGGATATCATCTGTGAGTACTTTGAGCCC 73  
 2Y 70 CAGCTCTACAGACAGGCTGCCTCATCTCTTGTCCCGGATGTGTAGTATCTGAG 129  
 2b 74 AAGCTCTCTCGAGAGGCTTGCTCATCTCTTGTCCAGCAAGATTGCATCTGTGTGAA 133  
 2Y 130 TTCTTACCATGGTCCAACTGTAGCAAGGATGTGGGAAGAAATGACAGCATAGAACTCGC 189  
 2b 134 TTTTCTGCTGTCGGAATCTCCAGAACCTGCGGACGGGCTCCAGCACCGACGCGT 193  
 2Y 190 CGGTCATAGTCCCTCTCTTGTGTGTGTGCAATGTCCAAATCTGACTAGTCAAGA 249  
 2b 194 CATGTGTGGCGCCCGGAGTTGGAGGCTGTGGTGTCTCCAACTGACGGAGTTCCAG 253  
 2Y 250 GCCTGTGATGCTCCATTTCTGTCTCTTGGGAAGAGGAATATACATTTAGCTTTAAG 309  
 2b 254 GTGTG-----CCAATCCAGTCCATGCGAGCGGCGAGAGCTCAGGTACAGCTGCAT 304  
 2Y 310 GTTGGACATGAGTAATGAGACTGCTCATCTTAAAGAAATTAATCCAGCGGAAGA 369  
 2b 305 GTGGGGCCCTGGAGCACCTGCTCAATGCCCACTCCCGACAGTAAGACAAGCAAGGAGA 364  
 2Y 370 ACTGTTCTGGAATTTAA-----CTCTGATTCAAATGAG 402  
 2b 365 CGCGGAAGAAATAAGAACGGGAAGAGACCGGACGAGAGGTAAAGGATCAGAGCC 424  
 2Y 403 CGAGTCACTTTAAACATCAAGTTACA-----AAGCACATCATCATTCGAAGTCT 453  
 2b 425 CGCGAGCTTTTAAAGAAAAGAGAAACAGAAAACAGGACAGAACAGAGAGACAATAAT 484  
 2Y 454 TGGCAATAGAGATAGTTATCAACCGGAGGTTTCGTGTACAGAGATGATGACAA 513  
 2b 485 TGGACATCCAGATTGGATATCAGACAGAGAGGTTATGTGCAATTACAGACGGGAAA 544  
 2Y 514 AATGCTATGTTAAGCCTTTGCTTTCAAGATTCTCTCCCATTTGACTGTTTCACTCTGCA 573  
 2b 545 GCTGCTGATTTAAGCTTTTGGCAGCAGAGAGAGCTTCCAATGACCTCCAGTCTCTGTG 604

574 ATGCCCAAGACTGTGAAACCTCCAGTGTCTCTCTGGAGCCCCCTGCTCCAGACATGC 633  
 575 ATGCCCAAGACTGTGAAACCTCCAGTGTCTCTCTGGAGCCCCCTGCTCCAGACATGC 634  
 605 ATACCAAAAGAGTCCAGGTTTCGAGTGTCTGAGTGGAGCCCCCTGCTCAAAAACATGC 664  
 634 CGTTCAGGAGTCTCTTCCAGGATTTAGGAGCAGGACCGGACCTGAGACACATGGCT 693  
 665 CATGACATGTGTCTCTGAGGACCTCGTGTAGGACACGAACCATCAGGAGTTTCCC 724  
 694 ATTGAGAGTGAAGAGGAGTGTCTTGAACCTCTTGAAGAAAGAGGCTCATTTGTTGAAGGA 753  
 725 ATTGCGAGTGAAGAGGAGTGTCCAGATTTTGAAGAAAGAAACCTGTTGTCTCAAGGA 784  
 754 GAATCTTGCAGCAATGTCCAGTATTTCTGAGAACTTCTGAGAACTTCTGAAATGAAAGATGCCAA 813  
 785 GATGAGTGTGCTCCCTGTGCCAGTATGGTGGAGAACTACAGAGTGGATGAGTGGCGT 844  
 814 GTCTCTCTCTCTCGAGCAGCAGGATCCCACTTGGCATGTGACGGACCCCTGTGTGGC 873  
 845 GTGAGACCTTTGCTCAGTCAGCAGGACAAAGAGCGCGGCAACAGACGGCCCTCTGTGGA 904  
 874 GGTGGGATCCAGACCCCGGAGGTGTACTGTGCCAG-----AGGTTACCAAGCA 921  
 905 GGGGCGCATCCAGACCCGAGAGGTGTACTGTGCTGAGGCCCAACGAAACCTCTCTCAAA 964  
 922 GCTCGGCACTGAGGGCCCAAGGAGTCTCTAGACCTGTGAAAGGCAATTATGTGTGCGGA 981  
 965 TTAAGTATCCCAAGAAACAAAGAGCCTCAAGCAATGAGCTTTAAATTTATGACATGGA 1024  
 982 CCCGCCCTGCTCTCAGCTCTGCAATATCCCTTGTCTACGGACTGCTAGTATCT 1041  
 1025 CCTATCCCTAATACTATACAGAGCTGTGCACATCTCTTGTCCAACTGAATGAAAGTTTCA 1084  
 1042 TCCTGTGAGCCTGGGCTGTGATCCATGAAACTCTGATGAACTCTCAGGCGGAAAGAA 1101  
 1085 CTTTGTGAGCTTGGGACCTTGTACTTATGAAACTGTATGATCAGCAAGGAAAGAA 1144  
 1102 GGATTTAGAACGAGCGCAGCG-----CCATGCTCTCATGGAATCTACAGGCGCTGCA 1152  
 1145 GGTTCAACTGAGGAGCGCGCATTAACCAATGAGCCACTGAGGCTCTGGGTTAAC 1204  
 1153 GGGCANTGCCCTCATTTGGTGGAGTGTCTTCTTGTGAGGATCCAAATGTGCTACGATGG 1212  
 1205 GAAACTGCCCTCACTTACTGGAAGACCATCTCCCTGTGAAGAGCTGCTGTATGACTGG 1264  
 1213 ---CTGGCATCAGAGGAGTCTGTTCCCTCATCATGAA---AATGTGGCTCGGACAT 1266  
 1265 AAGCGGTGAGACTGGGAGACTGAGCCAGATTAACGAAAGGAGTGTGTTCCAGGACG 1324  
 1267 CGTATTCTGAAGCGCTGTGCCAGAAAGACCGCGAGAGATGTATCAGGAGTCTTTGC 1326  
 1325 CAAAGTTCAAGAGGTTGTGTGCATCAACAGTGTGAGAGAAAGTTGACAGACAGCTGTGC 1384  
 1327 CCAGTTCCTCCCTCTCTGAGAGGAGTCTTGTGAAATTCCTCCGAAATGAGCTGTGTG 1386  
 1385 AGAGATGCCATCTTCCCATCTGAGGCTGTGATGCCCATATCCCGGAAAGATGTGTG 1444  
 1387 CTGACGAGTGAAGAGTGTGATCTCTGTTCCAGTCTCTGTTCAAAATFAAAACCTCAGAT 1446  
 1445 CTCACACATGCTTACGTTGTCTCTGCTCAGACACCTCTCAGGAGAAACGACAGAA 1504  
 1447 GGGAAACAGACAGGTCAAGAACTATCTGGC---ACTGGCTGGGAGAGTGGAAAGCA 1503  
 1505 GGGAAACAGATACAGAGCAGATCCATCTGCGCTATGCGGTTGAAGAGGTTGAATTCGC 1564  
 1504 TGTCCTCCCTTAGTCAGGCTCTCAAGAGCATGTTTGTGTATGACCATCTCTGTATGAG 1563  
 1565 TGTCCTCAATAGCAGTGTCTTGCAGAGATGAGAGTGTATGAGCATCTTGCACAGTG 1624  
 1564 CTTCTACTGGAGACATCCCTTGGGCGCTTGTGTGTGAGGACACATTGTGTAACCTGCTT 1623  
 1625 TACCACTGGCAAACTGGTCCCTGGGCGCAGTGTGATTTAGGACACCTCAGTATCTCTTC 1684  
 1624 AATGCAACCATTTGGCTGGAATGGAGAACCCAGCTGTGTGTAGGCAATTCAGACTCGGAGA 1683

1685 AACACAACTACGACTTGAATGGGAGGCTCTCTCTGTGCGCATGCACACAGAAA 1744  
 1684 GTCCTCTGTGTCAAGAGTCACTGTGGGCAAGTAATGACCAAAAGATGTCAGATTCTACT 1743  
 1745 GTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAATATGTCCTGAAGCCTT 1804  
 1744 CGACTGAACTGTGCGGCCCTGTCTTCCATGCAAAAAGACTGTATTGTCACTGCT 1803  
 1805 CGACTGAACTGTGAAGCCTGTCTGCTCTCTGTTGAAGAGGACTGTATTGTGACCCCA 1864  
 1804 TTCAGTGTAGTGACACCTGCGCCAGGATGTGCCAAGCAGGAATGCCACAGT---AAAA 1860  
 1865 TATAGTGTAGTGACATATGCGCCCTCTCTGTTGTAAGHAGGGGACTCCAGTATCAGGAAG 1924  
 1861 CAGTCTCGATCAGAAATCATATCACAAGACAGCCAAATGAGGCCAGGAATGCCAGAT 1920  
 1925 CAGTCTAGGCATCGGGTCAATTCAGTGTGCGCAGCCAAACGGGGCCGAGACTGCACAGAT 1984  
 1921 ACCTTATATGAGGAGAGTGTGAAGATGTTCTTGTGCTGTATATATCGTGGGAAG 1980  
 1985 CCCCTCTATGAAGAGAGCCTGTGAGGACCTCAAGCGTGCCAAAGCTACAGTGGGAAG 2044  
 1981 CCACGAAATGAGGCCCTTGATCTTGTAGTGCAGAGTCTGTCTGCGCAGGGAATTAACGGGC 2040  
 2045 ACTCACAATGGCGCAGATGCCAATTAGTCCCTTGGAGCGTGCACAAGCAGCCCTTGA 2104  
 2041 AGCAGTGAAGCCTGTGGAAGGGGTATACAAACAGAGCTGTCTCATGATCTGTGATGAC 2100  
 2105 GCACAGGAGGCTGTGGGCCCTGGGGCGACAGGCAAGAGCCATTACTTGTGCGCAAGCAAGAT 2164  
 2101 AACCGGTGACGAGAAATGATGGAATGCCCTCAAGCAGACAAACCGCATGCCCTCTCTTGTG 2160  
 2165 GGAGGACAGGCTGGAATCCATGAGTGCCCTACAGTATGCGAGCCCTGTGCGAGCCCTTACC 2224  
 2161 CAAGAAATGCAAGTCCCATGTCGAGAAAGCTGCACCTTCACTGTGCTGGTCCAAATTAAG 2220  
 2225 CAGGCTCCAGATGCCCTGCGCAGGATGACTGTCAAATTGACGAGCTGGTCCAAATTTCT 2284  
 2221 CCCTGCTCCAGAACTGTGAAGCCACAAATAAGTAGGGGCGACAGCTCACAGGGAAGAGC 2280  
 2285 TCATGCAATGAGACTGTGTGCGAGTATAGGACCAAGAACGCACTCTTGTGGAAAAAGT 2344  
 2281 AGAAAGAGGAAATGCCAGATTCGACCTTTACCTTTAGTGGAGACAGAACTATGT 2340  
 2345 AAAAAGAGGAAAAATGTAAAAATTTCCCATTTGTATCCCTGATGAGACTCAGTATTGT 2404  
 2341 CCTTGTATGATTTATATCCCACTTATGAACTGGTCAAGTTCATCTTCCAGAA 2400  
 2405 CTTGTGCAATATATATGCAACCTGTGGGAACTGGTCAAGACTGTATTTTACCAGAG 2464  
 2401 GGCAGAGGAGCCTCACCGAGGACTGCGGTACAAAGCAGACAGCAAGAAATGTGGAGAA 2460  
 2465 GGAAAGTGGAACTGTTGCTGGAAATGAAGTACAAAGAGACATCAAGGAATGCGGACAA 2524  
 2461 GGCCTGCGCTTTCGAGCAGTAGCCTGTCTGATAAAAATGAAACCTGTGACCCCTCC 2520  
 2525 GGATATCGTTTACCAAGCAATGGCATGTACCATCAAAATGCGAGCTGTGTGAAACATCT 2584  
 2521 TTCTGACAGCTCTGTTTACATTCAGAAAAATGTGTCTTCCCTGCGCCATTTGATTGC 2580  
 2585 AGATGTAACAGCCTATGTTACATTTGAGGAGGCTGCATCATCCCTGCCCTCAGACTGC 2644  
 2581 AGTTAAGGATGTGTAGTTGGGGTCTGAGTTCATCTTGTGAAATGGAGTGAAG 2640  
 2645 AAGCTCAGTGTGTGTCAACTGTGCGGCTGCAGCAAGTCTCTGTGGAGTGTGTGAAG 2704  
 2641 ATTTCGATCAAAATGCTTAAAGAAAACTTACAAATGAGGAGCAGCATGTCCTCAACTG 2700  
 2705 GTTCTCTTAATGCTGCTGGAAGAACCAATATATGAGGAGAGGCTTGGCCCCAACTG 2764  
 2701 GATC-----TCAAGAAATCAGGTACATGAGCAGTCCCATTTTACAGTGTGCAATCAG 2754

2765 GACCATGTCAACAGGACACAGGTGTATGAGTGTGTCCTCATGCCACAGTGAAGCAACCAG 2824  
 2755 TATTCTGGGTTGTAGAACACATGTTCTTTCATCAAAAACAAATGAGTGAAGTCCCTG 2814  
 2825 TACCTAAGGTTCAGAGAGCCCTGGAGCATCTCAAGTGTACCTTTGTAATATGGGGAG 2884  
 2815 CCGTGTGGAGGAGAACAAATCTAGGAAAATTCAGATGTGT---GAATACTCGGATGT 2871  
 2885 AACTGTGGAGAGGGGTGCAAAACCCGAAAAGTGAATGCATCAGAAATACAGCAGATGSC 2944  
 2872 GAAGGTGGAGCAGTGGATAGCAACCTGTCAACAGGATGAATTCCTCCAGAAAACCCAG 2931  
 2945 CCTTCTGAACATGTAGAGATTACCTCTGTGACCCAGAGAGATGCCCTGGGCTTAGA 3004  
 2932 TCTGTGTCTCTTATGTGTCTCCCAATGAGTGTGTCTGTGAGTGGGACTTTTGGAGCAAA 2991  
 3005 GTGTGCAAAATTACCATGCCCTGAGGACTGTGTGATATCTGAATGGGTCCATGGACCCAA 3064  
 2992 TGCCCCACAGTCAATGCAATCCCAACAAAT--GCAGAGAGAACTCGCCACCTGTAGA 3048  
 3065 TGTGTTTTGCCCTTGAATCAAAAGCAGTTTCCGGCAAGGTCACTGATCCCATCAGACAA 3124  
 3049 CCATCACTGAATCTCAAGGACTTGTGTGAAGACTCACAGGTGCAGCTTGCCTCTCTGAAT 3108  
 3125 CCAGCTGATGAAGAGAGATCTTGGCCCTAATGCTGTGAGAAAGAACCTGTAACTGAAAC 3184  
 3109 GAAATTTGCTTCCAGTTCAGTACAAATCTAAACAGAGTGGAGACATGCCAGCTGAGTAA 3168  
 3185 AAAAATGCTGTACCTATGATTAATGTAAACAGACTGGAGTACATGTGAGTGTGAG 3244  
 3169 AACGCAACCTGTGTGCAAGGCTCAGGACCCCGCTGAAGCTGTGTGAGTGTGATGATGSC 3228  
 3245 AAGGCAATTTGTGGAATATGAATATAAACAGAGATGTTGGATTGTGTTGAAAGTATGSC 3304  
 3229 AAGCCAGTCAAGTCAAGCAATGTGAGCAGCAATATTTGGAGAAACCCAGAGAAATGAGC 3288  
 3305 AAGTCAGTTGACTGAAATATATGTAAGCGCTTGGCTTGGAGAAACCTGGCAGATGAAC 3364  
 3289 ATTCCCTGCTTGGTGGATGCGTGTCAACTGTCACTGAGGTGAGCGCTTGGAGCA 3348  
 3365 AGTCTCTGATGAGTGTGATGCGCTGTGAATCTGAGCTTTCTGATGTCTCTTGGTCA 3424  
 3349 GAGTGTTCACAGACCTGTGGCATGGAGTGAATGAGCCGAGCTCGATTTATCATATTG 3408  
 3425 GAATGTTCTCAAACTATGTGGCTCACAGGAAAAATGATCCGAAAGCAACAGTGAAGCCAG 3484  
 3409 CCAACCCAGAGAGAGGAGCGGCATGCCCAACAGAGCTTACCCAGAGAAACCTGCCCA 3468  
 3485 CCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACCACTCCAAACCCCTGCCCA 3544  
 3469 GTGACCCCTGCTACAGCTGGGTCTTGGCAACTGCTGTGATGTAAATTTGAGGGTGA 3528  
 3545 GTGAAGCCTTGTATCGGTGGCAATATGGCCAGTGTCTCCATGCCAAGTGCAGAGGCC 3604  
 3529 GACTGTGGGAAAGAGTTCAGATCCGAGCCTTCTGAGCTGCTCCAGTGGTCCAGTGGTCAATA 3588  
 3605 CAGTGTGGAGAGGAGCCAGAAACAGGAAACATTTCTGTGTAGTAAATGATGGGTGAGT 3664  
 3589 TCTCATGCACTGGACGTGTGAGGATGCACTGTGTGAGAAATGCCCTT----- 3638  
 3665 GATGATTTTCAAGAAAGTGGTGTGATGAGAAATCTGTGCTGACATTTGAATCATATTAGAT 3724  
 3639 -----TCAGGACAGCATCTGGAAGCAGTGTGTGCTGCTTGGCCCTGCGCAGGAGACTGCCAT 3693  
 3725 GGTAATAAAATATGGTCTGTGAGGAATCTCTGAGCCAGCCTTGGCCAGGTGACTGTTAT 3784  
 3694 TTAAACAGAAATGTTCAGAGTGGAGCACATGTGAATTAACCTGCATTTGAGGAAGACTTT 3753  
 3785 TTGAAGACTGTGTCTTCCGAGGAGCTGTGTGAGTGTGCTGTGTAATGTTGAGGATCTA 3844  
 3754 GAGACTGTGGCGCCGAGTCTAGATCAAGGACTTTTATAATTCAGTCTTTTGGAGAACAA 3813  
 3845 GGTCTTGGTGAATACAGGTGAGTCCAGACCGGTGATTATACAAAGAACTAGAGAAATCAG 3904



QY 574 ATGCCAAAGACTGTGAACCTCCAGTGTGTCCTCTCGAGGCCCTGTCTCCAGACATGCG 633  
DB |||||  
1096 ATCACAAAGAGTGCACAGTTTCAGAGTGGTTCAGAGTGGAGCCCTGTCTCAAAACATGC 1155  
QY |||||  
634 CGTTCCAGGAGTCTCTTGCCAGGATTTAGGAGCAGGAGCCGGAACGTCGAAGCACATGGCT 693  
DB |||||  
1156 CATGACATGGTCCCTCCCTCAGGCACTCGTGTAGGACACGACCATCAGCAGTTTCCC 1215  
QY |||||  
694 ATTGAGGTGGAAGAGTGTCTCTGAATCTTTGAGAAAGAGGCTGTCAATTTGTAAGGA 753  
DB |||||  
1216 ATTGACAGTGAAGAGTGTCCAGATTTGAGAAAGAAAGAACCTGTCTGCTCAAGGA 1275  
QY |||||  
754 GAACTTCTGACAGATGCCCAGGTATCTCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
DB |||||  
1276 GATGGAGTTGTCCTCTGTGCCAGTATGGCTGGAGAACTACAGAGTGGACTGTAGTGCCGT 1335  
QY |||||  
814 GTCCTCTCTCTCTCGAGCAGCAGGATCCCCACTGGCATGTGAACGGACCCGTGTGGC 873  
DB |||||  
1336 GTGGACCCCTTTGCTCAGTCAGCAGGACCAAGAGGCGCGCAACAGACGGCCCTCTGTGGA 1395  
QY |||||  
874 GGTGGGATCCAGACCCGGAGGTACTGTGCCAG-----AGGTACCAGCA 921  
DB |||||  
1396 GGGGCACTCCAGACCCGAGAGGTGTACTGTGCGTGGAGCCCAACGAAACCTCTCTCAAA 1455  
QY |||||  
922 GTCGCCCACTGAGGCGCAAGGAAGTCTCTAGACCTGTGGAAGGCAATATGTGTGGA 981  
DB |||||  
1456 TTAAGTACCACAAGAACCAAGAGCCCTCAAGGCCAATGGACTTAAATATATGCAC-TGA 1515  
QY |||||  
982 CCGGCCCGTTCCTCTCAGTCTGGAATATCTCTGCTCTACGACTGCATGATATCT 1041  
DB |||||  
1516 CCAATCTCTTAATPACACAGCTGTGGCAGATTCCTTGTCCAACCTGATGTGAAGTTTCA 1575  
QY |||||  
1042 TCTGCTCAGCCTGGGCGCTGTGCATCATGAAACTGTGTAACCTCAGGCGGAAAGAA 1101  
DB |||||  
1576 CTTGCTCAGCTTGGGACCTTGACTTATGAAACTGTATGATCAGCAGGAGAAAGAA 1635  
QY |||||  
1102 GAATTTAGAAAGAGCAGCG-----CAATGCTCTCATGGAATCTPACAGGCGCTGCA 1152  
DB |||||  
1636 GGCCTTCAAACTGAGGAAGCGGCGCAATACCAATGAGCCCACTGGAGGCTCTGGGGTAAAC 1695  
QY |||||  
1153 GGGCATTCGCCCTCATTTGGTGGAGTCTGTCTCTGTGAGGATCCAAATGTGTCTACCGATGG 1212  
DB |||||  
1696 GGAATCGCCCTCACTTACTGGAAGCAATTCCTGTGGAAGCGCTGCCTGTATGACTGG 1755  
QY |||||  
1213 ---CTGGCATCAGAAGGATGTGTTCCCTGATCATGGAA---AATGTGSCCTGGGACAT 1266  
DB |||||  
1756 AAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAGGAGTGTGTCCAGGCACG 1815  
QY |||||  
1267 CGTATTTCTGAGGCGGTCTGCCAGATGACCGCGAGAGATGTATCAGGAGTCTTTGC 1326  
DB |||||  
1816 CAAAGTTCAAGAGTGTGTGCAATCAACAGTATGAGGAAGAGTTGACAGACAGCTGTGC 1875  
QY |||||  
1327 CCAGTTCCCCCTCTCTGAGAGGAAGTCTTGTGAAATTCCTCTGCCGAATGGAATGTGTG 1386  
DB |||||  
1876 AGAGATGCCATCTTCCCATCCCTGTGSCCTGTGATGCCCATGCECGAAGACTGTGTG 1395  
QY |||||  
1387 CTGAGCGAGTGAACGAGTGTCTATCTGTTCCAGTCTCTGTTCAATATAAAATCTCAGAT 1446  
DB |||||  
1936 CTCAGCAATGCTCTACGTGTCTCTCTGCTCACAACCTCTCTCAGGGAAGAACAGACGAA 1995  
QY |||||  
1447 GGGAAACAGACAGGCTCAAGAACTATCTCTGGC---ACTGGCTGGGGAAGGTGGAAAGCCA 1503  
DB |||||  
1996 GGGAAACAGATACAGACAGATCCATCTCTGGCCATATGCGGTGAGAGGTGGAAATTCG 2055  
QY |||||  
1504 TGTCCCTCTAGTCAGCTCTCAAGAGCATCGTTTGTGTAAATGACCAATCTCTGTATGACG 1563  
DB |||||  
2056 TGTCCAAATAGCAGTGTCTTTGCAAGAGTACGAAGCTGTAAATGAGCATCTCTTGCACAGT 2115  
QY |||||  
1564 CTTCACTGGGAGACATCGCTTGGGCGCTTGTCTGAGGACATTTGGTAACTGGCCCTT 1623  
DB |||||  
2116 TACCACTGGCAACTGTGTCTTGGGCGCGTGCATTTGAGGACACCTCAGTATGCTCTTC 2175

QY 1624 AATGCAACCAATTTGGCTGGGAATGGAGAGCCACGTTGTGTGTAGGCAATTCAGACTCGGAGA 1683  
DB |||||  
2176 AACACAACTACGACTTTGGAAATGGGAGGCTCTCTGCTCTGTGGCATGCAGACAAGAAA 2235  
QY |||||  
1684 GTCTTCTGTCTCAAGAGTCACGTGGGACAAAGTAATGACCAAAAGATGTCAGATTTCTACT 1743  
DB |||||  
2236 GTCACTCTGTGGAGTCAATGTGGCCAAAGTGGGACCCAAAATGTCTCTGAAGCCCTT 2295  
QY |||||  
1744 CGACCTGAAACTGTGGGCCCTCTGTTTCTCCATGCAAAAAAGACTGTATTTGTGCTGT 1803  
DB |||||  
2296 CGACCTGAAACTGTAGGCGCTTGTCTGCTCTCTTGTGTAGAGGACTGTATTTGTGACCCCA 2355  
QY |||||  
1804 TTCAGTGTGTGACACCTCTGCCCAAGGATGTCAGCAGCAGGAATGCGACACT---AABA 1860  
DB |||||  
2356 TATAGTGTGACATGACATCATGCCCCCTTCTGTGTAAAGAGGGGACTCCAGTATCAGGAAG 2415  
QY |||||  
1861 CAGTCTCGATACAGAAATCATATCAAGAAAGCAGCAATGGAGGCGCAGGAATGCCCAGAT 1920  
DB |||||  
2416 CAGTCTAGGCATCGGCTCATCATTCAGCTGCCAGCAACCGGGGCGGAGACTGCACAGAT 2475  
QY |||||  
1921 ACCTTATATGAGGAGAGAGTGTGAGATGTTTCTTGTGTCTGTATATCGGTGGAG 1980  
DB |||||  
2476 CCCCCTTATGAAGAGAGGCGCTGTGAGGCACCTCAAGCGTGCCAAAGCTTACAGGTGGAG 2535  
QY |||||  
1981 CCACAGAAATGGAGCCCTTGCATCTTAGTCCAGAGTCTGTCTGCGCAGGAATAACGGGC 2040  
DB |||||  
2536 ACTCACAATGGCGCAGATGCCAATTAGTCCCTTGAGCGGTGCAACAGACAGCCCTGGA 2595  
QY |||||  
2041 AGCAGTGAAGCCCTGTGGAAAGGGTTTCAAAACAGAGTGTCTCATGTATCTCTGATGAC 2100  
DB |||||  
2596 GCACAGAAAGGCTGTGGGCTCTGGCGACAGGCAAGAGCCATTTACTTGTGCGAAGCAAGAT 2655  
QY |||||  
2101 AACCGTCTCAGCAAAATGATGAATGCTCAAGCAGACAGAAACGGCATGCTCTCTCTTGTG 2160  
DB |||||  
2656 GAGGACAGCTGGATCCATGATGTGCTTACAGTATGAGGCGCTGTGCCAGCCCTTACC 2715  
QY |||||  
2161 CAAGAAATGACAGTCCCATGTCTGAGAGAGTGCACCTTCACTGTGTGGTTCGAAGTTTACG 2220  
DB |||||  
2716 CAGGCTTGCAGATCCCTCTGCGAGGATGACTGTCAATTCACAGCTGTGTCGAAGTTTCT 2775  
QY |||||  
2221 CCTGTCTCCACGAATGTGAAGCCACAAAAGTAGCGCGCAGACGCTCACAGGGAAGAC 2280  
DB |||||  
2776 TCAATGCAATGGAGACTGTGTGTGAGTGTAGGACCAAGAGCGCACTCTTGTGTGAAAAGT 2835  
QY |||||  
2281 AGAAGAAGGAGAAATGCCAGGATTTGACCTTTACCTTTAGTGGAGACAGAACTATGT 2340  
DB |||||  
2836 AAAAGAAGGAAATGTAAAAATTTCCATTTGTATCCCTGATTTGAGACTCAGTATTTGT 2895  
QY |||||  
2341 CCTGTGTGAATTTATATCCCAACCTTATGGAACCTGTGAGATTCATTTCTTCAGAA 2400  
DB |||||  
2896 CCTGTGTGCAAAATATAATGCAACCTGTGGGGAATGCTGTGAGACTGTATTTTACAGAG 2955  
QY |||||  
2401 GGCAGAAAGGAGCGCTCACCGAGGACTGCGGGTACAAGCAGACAGCAAAAGAAATGTGGAGAA 2460  
DB |||||  
2956 GGAAGATGGAAGTGTGTCTGGATGAAGTACAGGAGACATCAAGGAATTCGGACAA 3015  
QY |||||  
2461 GGCCTCGCTTTTCGAGCAGTAGCCTGTCTGTATAAAATGGAAGACCTGTGTGACCCCTCC 2520  
DB |||||  
3016 GGATATCGTTTACCAAGCAATGGCATGTACGATCAAAATGGCAGGCTGTGTGAAAACATCT 3075  
QY |||||  
2521 TTTCTGAGCAGCTGTGTTTACATTTCAAGAAAATGTGTTCATTTCCCTGCCCATTTTCAATTGC 2580  
DB |||||  
3076 AGATGTAAACAGCCATGGTTTACATTTGAGGCGCTGCATCCCCCTGCCCTCAGACTGC 3135  
QY |||||  
2581 AGTTAAGCAATGTGTCTAGTTGGGGTCTTTCAGTTTCACTTTGTGGAATTTGGAAGTGAGA 2640  
DB |||||  
3136 AAGCTCAGTGTAGTGTGCTCAACTGTGTGCGCTGTGCAAGAGTCTGTGTGGAGTGGTGTGAAG 3195  
QY |||||  
2641 ATTTCGATCCAAATGCTAAAGAAAACCTTACAAATGGAGGACGACCATGTCCTCAAACTG 2700  
DB |||||  
3196 GTTCGTTCTTAAATGGCTGGTGAAGAAACCATATAATGGAGGAAGGCGCTTGGCCCAACTG 3255  
QY |||||  
2701 GATC-----TCAAGAAATCAGGTATAGGCGAGTCCCATGTTTACAGTGTGCAATCAG 2754

```
3256 GACCATGTCAACAGGACAGAGTGTATGAGGTGTCCCATGCCACAGTGAACACAG 3315
3257 |
3258 |
3259 |
3260 |
3261 |
3262 |
3263 |
3264 |
3265 |
3266 |
3267 |
3268 |
3269 |
3270 |
3271 |
3272 |
3273 |
3274 |
3275 |
3276 |
3277 |
3278 |
3279 |
3280 |
3281 |
3282 |
3283 |
3284 |
3285 |
3286 |
3287 |
3288 |
3289 |
3290 |
3291 |
3292 |
3293 |
3294 |
3295 |
3296 |
3297 |
3298 |
3299 |
3300 |
3301 |
3302 |
3303 |
3304 |
3305 |
3306 |
3307 |
3308 |
3309 |
3310 |
3311 |
3312 |
3313 |
3314 |
3315 |
3316 |
3317 |
3318 |
3319 |
3320 |
3321 |
3322 |
3323 |
3324 |
3325 |
3326 |
3327 |
3328 |
3329 |
3330 |
3331 |
3332 |
3333 |
3334 |
3335 |
3336 |
3337 |
3338 |
3339 |
3340 |
3341 |
3342 |
3343 |
3344 |
3345 |
3346 |
3347 |
3348 |
3349 |
3350 |
3351 |
3352 |
3353 |
3354 |
3355 |
3356 |
3357 |
3358 |
3359 |
3360 |
3361 |
3362 |
3363 |
3364 |
3365 |
3366 |
3367 |
3368 |
3369 |
3370 |
3371 |
3372 |
3373 |
3374 |
3375 |
3376 |
3377 |
3378 |
3379 |
3380 |
3381 |
3382 |
3383 |
3384 |
3385 |
3386 |
3387 |
3388 |
3389 |
3390 |
3391 |
3392 |
3393 |
3394 |
3395 |
3396 |
3397 |
3398 |
3399 |
3400 |
3401 |
3402 |
3403 |
3404 |
3405 |
3406 |
3407 |
3408 |
3409 |
3410 |
3411 |
3412 |
3413 |
3414 |
3415 |
3416 |
3417 |
3418 |
3419 |
3420 |
3421 |
3422 |
3423 |
3424 |
3425 |
3426 |
3427 |
3428 |
3429 |
3430 |
3431 |
3432 |
3433 |
3434 |
3435 |
3436 |
3437 |
3438 |
3439 |
3440 |
3441 |
3442 |
3443 |
3444 |
3445 |
3446 |
3447 |
3448 |
3449 |
3450 |
3451 |
3452 |
3453 |
3454 |
3455 |
3456 |
3457 |
3458 |
3459 |
3460 |
3461 |
3462 |
3463 |
3464 |
3465 |
3466 |
3467 |
3468 |
3469 |
3470 |
3471 |
3472 |
3473 |
3474 |
3475 |
3476 |
3477 |
3478 |
3479 |
3480 |
3481 |
3482 |
3483 |
3484 |
3485 |
3486 |
3487 |
3488 |
3489 |
3490 |
3491 |
3492 |
3493 |
3494 |
3495 |
3496 |
3497 |
3498 |
3499 |
3500 |
3501 |
3502 |
3503 |
3504 |
3505 |
3506 |
3507 |
3508 |
3509 |
3510 |
3511 |
3512 |
3513 |
3514 |
3515 |
3516 |
3517 |
3518 |
3519 |
3520 |
3521 |
3522 |
3523 |
3524 |
3525 |
3526 |
3527 |
3528 |
3529 |
3530 |
3531 |
3532 |
3533 |
3534 |
3535 |
3536 |
3537 |
3538 |
3539 |
3540 |
3541 |
3542 |
3543 |
3544 |
3545 |
3546 |
3547 |
3548 |
3549 |
3550 |
3551 |
3552 |
3553 |
3554 |
3555 |
3556 |
3557 |
3558 |
3559 |
3560 |
3561 |
3562 |
3563 |
3564 |
3565 |
3566 |
3567 |
3568 |
3569 |
3570 |
3571 |
3572 |
3573 |
3574 |
3575 |
3576 |
3577 |
3578 |
3579 |
3580 |
3581 |
3582 |
3583 |
3584 |
3585 |
3586 |
3587 |
3588 |
3589 |
3590 |
3591 |
3592 |
3593 |
3594 |
3595 |
3596 |
3597 |
3598 |
3599 |
3600 |
3601 |
3602 |
3603 |
3604 |
3605 |
3606 |
3607 |
3608 |
3609 |
3610 |
3611 |
3612 |
3613 |
3614 |
3615 |
3616 |
3617 |
3618 |
3619 |
3620 |
3621 |
3622 |
3623 |
3624 |
3625 |
3626 |
3627 |
3628 |
3629 |
3630 |
3631 |
3632 |
3633 |
3634 |
3635 |
3636 |
3637 |
3638 |
3639 |
3640 |
3641 |
3642 |
3643 |
3644 |
3645 |
3646 |
3647 |
3648 |
3649 |
3650 |
3651 |
3652 |
3653 |
3654 |
3655 |
3656 |
3657 |
3658 |
3659 |
3660 |
3661 |
3662 |
3663 |
3664 |
3665 |
3666 |
3667 |
3668 |
3669 |
3670 |
3671 |
3672 |
3673 |
3674 |
3675 |
3676 |
3677 |
3678 |
3679 |
3680 |
3681 |
3682 |
3683 |
3684 |
3685 |
3686 |
3687 |
3688 |
3689 |
3690 |
3691 |
3692 |
3693 |
3694 |
3695 |
3696 |
3697 |
3698 |
3699 |
3700 |
3701 |
3702 |
3703 |
3704 |
3705 |
3706 |
3707 |
3708 |
3709 |
3710 |
3711 |
3712 |
3713 |
3714 |
3715 |
3716 |
3717 |
3718 |
3719 |
3720 |
3721 |
3722 |
3723 |
3724 |
3725 |
3726 |
3727 |
3728 |
3729 |
3730 |
3731 |
3732 |
3733 |
3734 |
3735 |
3736 |
3737 |
3738 |
3739 |
3740 |
3741 |
3742 |
3743 |
3744 |
3745 |
3746 |
3747 |
3748 |
3749 |
3750 |
3751 |
3752 |
3753 |
3754 |
3755 |
3756 |
3757 |
3758 |
3759 |
3760 |
3761 |
3762 |
3763 |
3764 |
3765 |
3766 |
3767 |
3768 |
3769 |
3770 |
3771 |
3772 |
3773 |
3774 |
3775 |
3776 |
3777 |
3778 |
3779 |
3780 |
3781 |
3782 |
3783 |
3784 |
3785 |
3786 |
3787 |
3788 |
3789 |
3790 |
3791 |
3792 |
3793 |
3794 |
3795 |
3796 |
3797 |
3798 |
3799 |
3800 |
3801 |
3802 |
3803 |
3804 |
3805 |
3806 |
3807 |
3808 |
3809 |
3810 |
3811 |
3812 |
3813 |
3814 |
3815 |
3816 |
3817 |
3818 |
3819 |
3820 |
3821 |
3822 |
3823 |
3824 |
3825 |
3826 |
3827 |
3828 |
3829 |
3830 |
3831 |
3832 |
3833 |
3834 |
3835 |
3836 |
3837 |
3838 |
3839 |
3840 |
3841 |
3842 |
3843 |
3844 |
3845 |
3846 |
3847 |
3848 |
3849 |
3850 |
3851 |
3852 |
3853 |
3854 |
3855 |
3856 |
3857 |
3858 |
3859 |
3860 |
3861 |
3862 |
3863 |
3864 |
3865 |
3866 |
3867 |
3868 |
3869 |
3870 |
3871 |
3872 |
3873 |
3874 |
3875 |
3876 |
3877 |
3878 |
3879 |
3880 |
3881 |
3882 |
3883 |
3884 |
3885 |
3886 |
3887 |
3888 |
3889 |
3890 |
3891 |
3892 |
3893 |
3894 |
3895 |
3896 |
3897 |
3898 |
3899 |
3900 |
3901 |
3902 |
3903 |
3904 |
3905 |
3906 |
3907 |
3908 |
3909 |
3910 |
3911 |
3912 |
3913 |
3914 |
3915 |
3916 |
3917 |
3918 |
3919 |
3920 |
3921 |
3922 |
3923 |
3924 |
3925 |
3926 |
3927 |
3928 |
3929 |
3930 |
3931 |
3932 |
3933 |
3934 |
3935 |
3936 |
3937 |
3938 |
3939 |
3940 |
3941 |
3942 |
3943 |
3944 |
3945 |
3946 |
3947 |
3948 |
3949 |
3950 |
3951 |
3952 |
3953 |
3954 |
3955 |
3956 |
3957 |
3958 |
3959 |
3960 |
3961 |
3962 |
3963 |
3964 |
3965 |
3966 |
3967 |
3968 |
3969 |
3970 |
3971 |
3972 |
3973 |
3974 |
3975 |
3976 |
3977 |
3978 |
3979 |
3980 |
3981 |
3982 |
3983 |
3984 |
3985 |
3986 |
3987 |
3988 |
3989 |
3990 |
3991 |
3992 |
3993 |
3994 |
3995 |
3996 |
3997 |
3998 |
3999 |
4000 |
```

```
Db 4336 GGCTTTGGTGGAAATACAGGTCAGATCCAGACCGGTGATTATACAAGAACTAGAGATCAG 4395
QY 3814 GACAGCTGCCGCCCAACAGGTTCTAGAAACAGGCCCTTGTACAGAGGCAATGTTATCAC 3873
Db 4396 CATCTGTGCCCAAGCAGATGTTAGAAACAAATCATGTTATGATGGACAGTGTATGAA 4455
QY 3874 TACACATGGAAACCAAGCTCTTTGGAAACAATAAGCAACGAACCTGTATGTTGCCAGCGTTCA 3933
Db 4456 TATAAATGGATGGCCAGTCTTGGAGGGCTCTTCCGAAACAGTGTGTTCTCAAGGTCA 4515
QY 3934 GATGGCCTTAATGTACAGAGGCTGCTCCCTCAGGCCCGTCTGCTGTCATTCGGCAG 3993
Db 4516 GATGGTAAATGTAAACAGGGGGCTGCTTGGTATGAGCCAGCTGATGCCGACAGTCT 4575
QY 3994 TGCAATTCAGCCTGCAGAAACCTTTCTCTACTGTACACAGGGTGGAGTCTGTGGTGTGT 4053
Db 4576 TGTAACCCACCGTGTAGTCAACCCACTCGTACTGTAGCCGAGACAAAACATGCCATTGT 4635
QY 4054 GAGAAGGCTATACAGAGATAATGAATCAATGATGTTTCTGGAATTAATCTGCA 4105
Db 4636 GAAGAGGGGTACACTGAAGTCAATGCTCTTAAACAGCACCCCTTGAGCAATGCACACTTATC 4695
QY 4106 -----TGAAGTACCAGGCTCAGAGGATAAAGAGCTGATGTGAAACCTTTCTG 4157
Db 4696 CCCGTGTGTGTTATTTACCCACATGGAGGACAAAGAGGAGATGTGAAACAGTCCGGCT 4755
QY 4158 GAAAAACAGACCTGTGTAATCAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215
Db 4756 GTACATCAACCCAAACCTCCAGTAAACCCAGCAGGACGGGGAAGGACCTGGTTCTACAG 4815
QY 4216 CCATTGATCCAGATGSCCGAGTAAATTTGGGTTTATGGCGTTTCAGGTGGCGCTTTT 4275
Db 4816 CCATTGGGCGAGATGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTT 4875
QY 4276 CTCACTATGATTTTCTTAATTTACTTCTCTACTGTTTTCAGGAAGCA--AAAACA 4332
Db 4876 GTGTTACTCACTTTATTTGTTCTCATGATTTTCTAGCTTGCAGAAAGCCCAAGAACCC 4395
QY 4333 CATCAAGACACCTCCCAACAGAGCCTCTGACCTTACCTAGCTAGAGATTAGAC 4392
Db 4936 CAAAGAGGCAAAACACACCGACTGAAACCTTTAACTTACCTTATGATGAGATGCCGAC 4995
QY 4393 ATGTAA 4398
Db 4996 ATGTAA 5001

RESULT 7
AR338841 LOCUS 6378 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 332 from patent US 6569662.
ACCESSION AR338841
VERSION AR338841.1 GI:33725698
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 6378)
AUTHORS Tang,Y.T., Zhou,P. and Dmanac,R.T.
TITLE Nucleic acids and polypeptides
JOURNAL Patent: US 6569662-A 332 27-MAY-2003;
FEATURES Location/Qualifiers
source 1. 6378
/mol_type="genomic DNA"

ORIGIN
Query Match 28.0%; Score 1229.6; DB 6; Length 6378;
Best Local Similarity 57.6%; Pred. No. 0;
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;
QY 10 TGCAATTCAGAGCTGAACCGAGTGTGTTGCAATGAATATGCAACACTTGGCTT 69
```

Db 541 TGCATCCAGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 600  
Qy 70 CAGCCTCCTACAGAACAGCTTGCTCATCTCTTGTCCCGGATTTGTGTAGTATCTGAG 129  
Db 601 AAGCCTCTCCCTGGAGCAGCTTGCTCATCTCTTGTCCAGCAAGATTGCATCGTGTCTGAA 660  
Qy 130 TTTCTTACCATGCTCCACTGTAGCAAGGATGTGGGAAGAAATGGCAGCATAGACTCGC 189  
Db 661 TTTTCTGCTGCTCCGAATGCTCAAGACCTGCGGAGCGGCTCCAGCACCGGACGCGT 720  
Qy 190 GCGGTCTAGTCCCTCTCTTTTGGTGGTTTGAATGTCCAAATCTGACTGAGTCAAGA 249  
Db 721 CATGTGGTGGCGCCCGCCAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGAGTTCAG 780  
Qy 250 GCCTGTGATGCTCCCATTTCTGTCTCTTGGGAAGAGGATATACATTTAGCTTTAG 309  
Db 781 GTGTG-----CCAAATCCAGTCCATGTGAGGCCGAGGCTCAGGTACAGCCTGCAT 831  
Qy 310 GTTGGAACCATGGATTAATGACAGTCTGCTCATCTTAAAGAAATTAATCCAAAGCGGA 369  
Db 832 GTGGGCCCTGGAGCACCTGCTCATGTCCCACTCCCGACAGTAAAGCAAGCAAGGAGA 891  
Qy 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAAAATGAG 402  
Db 892 CGCGGGAAGATAAGAAACGGAAGGACCGCAGCAAGAGGATTAAGGATCCAGAGCC 951  
Qy 403 CGAGTCACTTTAAACATCAAGTTACA-----AAGCACATCATCTTTCGAGTCT 453  
Db 952 CGCAGCTTATTAGAAAGAGAGAAACAGAAACAGGACAGACAGAGCAAGACAANAAT 1011  
Qy 454 TGGCAATPAGAGATAGGTATCAAAACCGGAGGTTTCGTGTACAAGAGTGTAGACAA 513  
Db 1012 TGGACATCCAGATTGATATCAGACAGAGAGGTTATGTGCTAATCAAGACGGGAAA 1071  
Qy 514 AATGCTATGTTAAGCCTTTGCTTCAAGATTCCTTCCATTTGACTGTTTCAGTCTGCATC 573  
Db 1072 GCTGCTGATTTAAGCTTTTGGCAGCAGAGAGGTTCCAAATGACCTTCCAGTCTCTGTG 1131  
Qy 574 ATGCCCAAGACTGTGAACCTCCAGTGTCTCTTGGAGCCCTGCTCCAAAGACATGC 633  
Db 1132 ATCCAAAGAGTCCAGGTTTCCGAGTGTGTAGAGTGGAGCCCTGCTCAAAACATGC 1191  
Qy 634 CTTTCAGGAGTCTCTTGGCAGGATTTAGAGAGGAGGACCGGAACGTGAAGCAGATGGT 693  
Db 1192 CATGACATGTGTCCCTGCAAGGACCTCGTGTAAAGACACGAACCATCAGGAGTTTCCC 1251  
Qy 694 ATTGGAGTGAAGAGGAGTGTCTGAACTTCTCAGAAAGAGGCTCGATTGTTGAAGA 753  
Db 1252 ATTGGAGTGAAGAGGAGTGTCCAGATTTGAAGAAAGAGAACCTGTGTTGTCTCAGGA 1311  
Qy 754 GAACTTTGACAGCAATGTCCAGGTATTCCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 1312 GATGGAGTGTCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGGCGT 1371  
Qy 814 GTCTCTCTCTCTCGAGCAGCAGGATNCCCACTGGCATGTGCGGACCCGTTGTGGC 873  
Db 1372 GTGACCCCTTTGCTCAGTACAGAGCAAGAGGCGCGCAACAGACGCGCCCTCTGTGGA 1431  
Qy 874 GGTGGGATCCAGACCGGAGGTTACTGTGCCAG-----AGCGTACCAGCA 921  
Db 1432 GGGGCATCCAGACCGGAGAGTGTACTGGTGCAGGCCAACGAAACCTCTCTCAAA 1491  
Qy 922 GCTGCCGACTGAGGCGCAAGAGTCTCTAGACCTGTGGAAGGCAATTTATGTGGGA 981  
Db 1492 TTAAGTACCCCAAGAAACAAAGAGGCTCAAGGCAATGAGCTTTAAATTTATGCACCTGA 1551  
Qy 982 CCCGCCCGTTCCTCTCAGCTCTGCAATATCCCTTGTCTCAGGACTGCATGATATCT 1041  
Db 1552 CCTATCCCTATATACACAGCTGTGCCAATCTCTGTCCAATGATGTGAGTTTCA 1611  
Qy 1042 TCCTGGTCAGCTGGGCTGTGCATCCATGAAACTGTGCATGAACTCTCAGGGGAAAAA 1101  
Db 1612 CTTTGGTCAGCTGGGACCTTGTACTTATGAAACTGTAAATGATGATCAGCAAGGAAAAA 1671

Qy 1102 GGATTTAGAACGAGCAGCG-----CCATGTCCTCATGGAATCTACAGGCGCTGCA 1152  
Db 1672 GGCTTCAAACTGAGGAAGCGGCGCATTACCAATGAGCCACTGAGGCTCTGGGTAACC 1731  
Qy 1153 GGGCATTCGCTCANTTGGTGAGTCTGTTCTTGTGAGGATCCAAATGTCTACCGATGG 1212  
Db 1732 GGAAACTGCGCTTCACTTACTGGAAGCCATCCCTGTGAAGAGCCTGCTGTATGACTGG 1791  
Qy 1213 CTGGCA---TCAGAAAGGATCTGTTTCCCTGATCATGGA---AATGTGGCTTGGACAT 1266  
Db 1792 AAAGCAGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAGAGTGTGGTCCAGCAG 1851  
Qy 1267 CGTATTCTGAAGCCGCTGTCCAGAAATGACCGGAGAGATGTATCAGGAGTCTTTGC 1326  
Db 1852 CAAGTTCAGAGGTTGTGTGCATCAACAGTGTGAGAGAAAGTTGACAGACAGCTGTGC 1911  
Qy 1327 CCAGTTCGCCCTCTCTGAGAGGAAGTCTTGTGAATTCCTGCGGAATGGACTGTGTG 1386  
Db 1912 AGAGATGCCATCTTCCCATCCCTGTGSCCTGTGATGCCCATGCCCGAAGACTGTGTG 1971  
Qy 1387 CTGACGAGTGAACGAGTGTCTATCTTGTTCCTGAGTCTCTTCAAAATAAAATCTCAGAT 1446  
Db 1972 CTCAGCACATGCTTACGTGTCTCTCTCTCTCAGACACTCTCAGGGAACACAGAA 2031  
Qy 1447 GGGAAACAGACCAGCTCAAGAACTATCTCTGCG---ACTGGCTGGGAGGTGGAAAGCCA 1503  
Db 2032 GGGAAACAGATACGAGCAGATCCATCTGCGCTATGCGGTGAAGAGGTGGAAATTCG 2091  
Qy 1504 TGTCTCCCTAGTCTCAGAGCTCTCCAGAGCATGTTTGTGTAAATGACCATTCCTGTATGAG 1563  
Db 2092 TGTCCAAATAGCAGTCTTTCAGAAAGTACGAAGCTGTATGAGCATCTTGTGCACAGT 2151  
Qy 1564 CTTGCTGGAGACATCCCTTGGGCGCTTGTCTGAGGACATCTGTAACCTGCGCTT 1623  
Db 2152 TACCACTGGCAAACTGGTCTCTGGGCGCAGTGCATGTAGGACACTCAGTATCGTCTTC 2211  
Qy 1624 AATGCAACCATTTGGCTGGAATGGGAAGCCAGCTGTGTGTAGGCAATTCACACTCGGAGA 1683  
Db 2212 AACACACTAGACTTGGAATGGGAGGCTCTCTGCTCGCATGCACAGCAAGAAA 2271  
Qy 1684 GTCTTCTGTGTCAAGAGTCACTGTGGGACAAATGACCAAAAGATGTCAGATTTCTACT 1743  
Db 2272 GTCATCTGTGTGCGAGTCAATGTGGGCAAGTGGGACCCAAAAAATGCTCTGAAAGCCTT 2331  
Qy 1744 CGACTGTGAACTGTGCGCCCTGTTTCTCCATGCAAAAGAAAGACTGTATTGTGACTGCT 1803  
Db 2332 CGACTGTGAACTGTGAGGCTTGTCTGCTTCTGTGTAAGAGGACTGTATTGTGACCCCA 2391  
Qy 1804 TTCAGTGTGACACCTCGCCCAAGGATGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
Db 2392 TATAGTACTGGACATCATGCGCCCTCTTCGTGTAAAGAAAGGGGACTCCAGTATCAGGA 2451  
Qy 1861 CAGTCTCGATACAGAACTCATCTCCAGAACGCCATGGAGCCAGGATGCCAGAT 1920  
Db 2452 CAGTCTAGGCTCGGCTCATCTTCAGTGCAGCAACCGGGGCGGAGCTGCACAGAT 2511  
Qy 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTCTTGTGTCTCTGTATATCGGTGGAAG 1980  
Db 2512 CCCCTCTATGAAGAGAGGCTGTGAGSCACTCAAGCTGCCAAGCTACAGGTGGAAG 2571  
Qy 1981 CCACAGAAATGAGCCCTTGCATCTTAGTGCAGAGTCTGTCTGGCAGGGAATACCGGC 2040  
Db 2572 ACTCAAAATGGCGCAGATGCCAATTAGTCCCTTGGAGCGTGCACAGACAGCCCTGGA 2631  
Qy 2041 AGCAGTGAAGCCTGTGGAAAGGGTTACAAACAGAGCTGTCTCATGCTCTCTGTATGAC 2100  
Db 2632 GCACAGGAAGGCTGTGGGCTTGGGCGACAGCAGAGGCAATTTCTTGTCCGCAAGCAGAT 2691  
Qy 2101 AACCGGTGACGAGAAATGATGGAATGCTCAAGCAGACAAACGGCATGCCCTCTCTTGTG 2160  
Db 2692 GGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGCAAGGCGCTGTGCCAGCCCTTACC 2751

QY 2161 CAGAAATGCACAGTCCCATGTCGAGAGACTGCACCTTCACTGCTGGTCCAAAGTTACG 2220  
Db 2752 CAGCCCTGCCAGATCCCTCGCCAGATGATGTCATATGACCAAGCTGGTCCAAAGTTTCT 2811  
QY 2221 CCCTGCTCCACGAACTGTGAAGCCACAAAAAGTAGCGGCGACAGCTCACAGAGAAAAGC 2280  
Db 2812 TCATGCAATGGAGACTGTGTGTGCAATTTAGACACAGAAAGCGCACTCTTGTGTGAAAAGT 2871  
QY 2281 AGAAGAGAGAGAAATGCCAGGATTCGACTTACCTTTACCTCTAGTGAGACAGAACTATGT 2340  
Db 2872 AAAAAAATAAATAATGAAAAATTCCTATTTGTATCCCTGATGAGACTCAGTATGT 2931  
QY 2341 CCCTGTGATGAATTTATATCCCAACCTTATGAAAACTGTGTCAGATTGCATTTCTCCAGAA 2400  
Db 2932 CCCTGTGACAAATATATGACAACTGTGGGAACTGTGTCAGACTGTATTTTACAGAG 2991  
QY 2401 GGCAGAGGAGGCCCTCACCGAGGACTCGGGTACAAGACAGACAGCAAGAAATGTGAGAA 2460  
Db 2992 GGAAGAGTGAAGTGTTCCTGGGAAATGAAAGTACAAGGAGACATCAAGGAATCGGACAA 3051  
QY 2461 GGCCTGCGCTTCGAGCAGTAGCTGTCTGTATATAAATGGAAGACCTGTGACCCCTCC 2520  
Db 3052 GGATATCGTTTACCAAGCAATGGCATGCTACGATCAAAATGGCAGGCTGTGGAAAATCT 3111  
QY 2521 TTCTGCAGCAGCTCTGCTTACATTCAGAAAAATGTGTCTATTCCTGCCCAATTTGATTGC 2580  
Db 3112 AGATGTAACAGCCATGTTACATTTAGAGGCGCTGCATCATCCCTGCCCTCAGACTGC 3171  
QY 2581 AAGTTAAGCAATGGCTAGTGGGGTCTTGGAGTTCAGATTCATCTTGTGGAATGAGGTGAGA 2640  
Db 3172 AAGCTCAGTGAAGTGTTCCTGGGAAATGAAAGTACAAGGAGACATCAAGGAATCGGACAA 3231  
QY 2641 ATTCCGATCCAAATGGCTTAAAGAAAAACCTTACAATGGAGGAGACCATGTCGCCAACTG 2700  
Db 3232 GTTCGTTCTAAATGGCTGCGTGAATAAACCAATATATGGAGGAAGGCTTGGCCCAACTG 3291  
QY 2701 GATC-----TCAGAAATCAGGTACATGAGGCACTGCCATGTTTACAGTGAAGTCAATCAG 2754  
Db 3292 GACCATGTCAACAGGCACAGGTGTATGAGTGTGCCATGCCACAGTGAATGCAACACAG 3351  
QY 2755 TATTCCTGGCTGTAGAACACTGGTCTTCATGCAAAATCAACAATGAGCTGAGTCCCTG 2814  
Db 3352 TACCTATGGGTACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAG 3411  
QY 2815 CGCTGTGGAGGAGAAACAAATCTAGAAAAATCAGATGTGT---GAATCTGCGGATGT 2871  
Db 3412 AACTGTGGAGAGGCGTGCAACCCGAAAGTGAATGCAATGCAATACAGAAATACAGAGTGC 3471  
QY 2872 GAAGGTGGAGCAGTGATAGCAACCTGTGCAACCCAGGATGAATTTCCCGAGAAACCCAG 2931  
Db 3472 CCTTCTGAACATGTAGAGGATTAACCTGTGTGACCCAGAGAGATGCCCCCTGGGCTCTAGA 3531  
QY 2932 TCCTGTTCTCTTATGTGTCCTCAATGATGTGTCATGCTGAGTGGGAGCTTTTGGAGCAA 2991  
Db 3532 GTGTGCAAAATACATGCCCTGAGGACTGTGTGATATCTGATGGGGTCCATGGACCCAA 3591  
QY 2992 TGCCCAAGCTATCGCATCCCAACAAT---GCAGAGAGAACTGCCACCTGTGTAAGA 3048  
Db 3592 TGTGTTTGGCTTTGCAATCAAGACAGTTTCCGCAAAAGGTCACTGATCCCATCAGACAA 3651  
QY 3049 CCATCACTGAACTCAAGGACTGTGTGTAAGACTCACAGTGCAGGCTTGCCTCTGTAAT 3108  
Db 3652 CCAGCTGATGAAGGAAGATCTTCCCTTAATGTCTGTTGAGAAAAGAACCTGTAACTGTAA 3711  
QY 3109 GAAAAATGCTTCCAGTTCCAGTCAATCTAACAGAGTGGAGCACATGCCAGCTGAGTGAA 3168  
Db 3712 AAAAACTGTACCACTATGATTAATGTAACAGACTGAGTACATGTCACTGTGAGTGAG 3771  
QY 3169 AACGCACTGTGGTCAAGCGCTCAGGACCCGCTGCTAAGCTGTGTGTCAGATGATGCC 3228  
Db 3772 AAGGCAGTTTGTGGAATGGAATATAAACAAGGATGTGGATTGTGTTTCAAGATGATGCC 3831  
QY 3229 AAGCCAGTCAGCATGGACCAATGTGAGGACGATTAATTTGGAGAGGCCCCCAGAGATGAGC 3288

Db 3832 AAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAAAGAACTGGCAGATGAAC 3891  
QY 3289 ATTCCCTGCTTGGTGGAAATGCGTGTCAACTGTAGCTCTCAGGCTTGGAGCGGCTTGGACA 3348  
Db 3892 AGCTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3951  
QY 3349 GAGTGTTCACAGACCTGTGGCCATGAGAGTTCGAATGAGCCGAGCTCGAATTTATCATTTAG 3408  
Db 3952 GAATGTTCTCAACATGTGGCTTCACAGAAAAATGATCCGAGACGAAACAGTGAACCCAG 4011  
QY 3409 CCAACCCCAAGAGAGAGGAGCGCCATGCCACAGAGCTTACCCAGAGAAACCTGCCCA 3468  
Db 4012 CCCTTTCAAGGTGATGAGAGACCATGCCCTTCCCTGATGACCACTCCAAACCTGCCCA 4071  
QY 3469 GTGACCCCTGCTACAGCTGGGTCTTGGCAACTGGTTCGATGTAAATTTGAGGGTGA 3528  
Db 4072 GTGAAGCCCTGTTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTCAGAGGCC 4131  
QY 3529 GACTGTGGGAGAGAGTTCAGATCCGAGCCCTTCCCTGCACTGGTCCACAGTGGTTCATA 3588  
Db 4132 CAGTGTGGAGAGGAGCCAGAAACAGGAACATTTCTTGTGTAGTAAGTATGGGTGAGCT 4191  
QY 3589 TCTCATGCACTGACGCTGTCAGAGTGCACCTGTGTGAGAAATGCCCT----- 3638  
Db 4192 GATGATTTTCAGAAAGTGTGTGATGAGGAATTTCTGTGCTGACATTTGAATTTATAGAT 4251  
QY 3639 -----TCAGGACAGCATCTCTGAAGAGCTGTGTCTGTGCTGCCCTGCCAGGAGACTGCCAT 3693  
Db 4252 GGTATAAAAAATATGTTCTGTGAGGAATCTCTGAGCAGCCCTTCCCGAGGTGACTGTAT 4311  
QY 3694 TTACAGAAATCGTCAGAGTGGAGCACATGTGAATTAACCTGCATTGATGGAAGAAGCTTT 3753  
Db 4312 TTGAAGGACTGTCTTCTCTGAGCCTGTGTGAGCTGACCTGTGTGAATGTGAGGATCTA 4371  
QY 3754 GAGACTGTGGGCCCCCACTGTAGATCAAGGACTTTTATAATTCAGTCTTTTGAAGAACAA 3813  
Db 4372 GGTCTTGTGTGAATACAGGTTCAGATCCAGACCGGTGATTTATACAAAGAACTAGAGATCAG 4431  
QY 3814 GACAGCTGCCCCCAACAGGTCTTAGAAACAGCCCTTGTACAGGAGGCAATGTTATCAC 3873  
Db 4432 CATCTGTGCCCAGAGCAGATGTTAGAAACAAATATCATGTTATGATGACAGCTGTCTGAA 4491  
QY 3874 TACATCGAAAGCAAGTCTTTGGAACAATAAGAACTGTATGTCGTCAGCGTTCA 3933  
Db 4492 TATAAATGATGAGCGAGTGTCTTGAAGGGCTCTTCCCGAACAGTGTGTGTCAAGGTCA 4551  
QY 3934 GATGCGTTAATGTACAGAGGCTGTCTCCCTCAGGCCCTGCTGCTGCTGCTGCTGCTGCTGCT 3993  
Db 4552 GATGCTATAAATGTAAACAGGGGGCTGTCTTGTGTATGAGCCAGCTGATGCCAGAGTCT 4611  
QY 3994 TGCAATCCAGCTTCAGAAACCTTTCTCTCTACTGTACACAGGCTGGAGTCTGTGTTGT 4053  
Db 4612 TGTAAACCAACCGT 4671  
QY 4054 GAGAAGGCTATACAGAGATAATGAATCAAAATGAAATGAAATGAAATGAAATGAAATGAAATGAA 4105  
Db 4672 GAAGAAGGCTACACTGAATCATGCTCTTAAACAGCACCTTGGAGCAATGCACACTTATC 4731  
QY 4106 -----TGAAAGTACAGGCTCAGAGGATTAAGAACTGATGTGAAAAACCTTTCTG-G 4157  
Db 4732 CCCGT 4791  
QY 4158 GAAAAACAGACCTGTGAATTTCAAAAAATACATGATATTTTAAAGGA--TGTCTCTCTCAA 4215  
Db 4792 GTACATCAACCCCAACCTCCAGTAAACCCAGGAGCGGAGAGGAGCTGTGTGTGTGTGTGTGT 4851  
QY 4216 CCATTTGATCCAGATGCGGAGTAAAAATTTGGGTTTATGGCGTTTCAGGTGGCGCTTTT 4275  
Db 4852 CCATTTGGCGCAGATGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTT 4911  
QY 4276 CTCATCATGATTTTCTTAATATTTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4332



Db 4312 GTGTTACTCATCTTTATTTGTTCTCCATGATTANTAGCTTGCAAAAAGCCAAAGAAACCC 4971  
Qy 4333 CATCAAGACACACCTCCCAACAGAGCCTCTGACCTTAGCTACGATGAGACTTAGAC 4392  
Db 4972 CAAAGAGGCAAAACACCGACTGAACCTTTAAACCTTAGCTATGATGAGATGCCGAC 5031  
Qy 4393 ATGTAA 4398  
Db 5032 ATGTAA 5037

RESULT 8  
AX818221  
LOCUS AX818221 5032 bp DNA linear PAT 10-DEC-2003  
DEFINITION Sequence 92 from Patent WO03068268.  
ACCESSION AX818221  
VERSION AX818221.1 GI:39723273  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE  
AUTHORS Ek,S., Borrebaeck,C.A. and Ehinger,M.  
TITLE Treatment&comma; diagnosis and imaging of disease  
JOURNAL Patent: WO 03068268-A 92 21-AUG-2003;  
BIOINVENT INTERNATIONAL AB (SE)  
FEATURES  
source Location/Qualifiers  
1..5032  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 25.7%; Score 1129.6; DB 6; Length 5032;  
Best Local Similarity 58.5%; Pred. No. 0;  
Matches 2265; Conservative 0; Mismatches 1524; Indels 81; Gaps 14;  
Qy 610 TGGAGCCCTCTCTCAAGACATGCGCTTCAGGAGTCTCTGCGAGATTTAGGAGCAGG 669  
Db 4 TGGAGCCCTCTCAAAACATGCCATGATGCTGCTCCCTCGAGCATCTGTAAAGG 63  
Qy 670 AGCCGGAACGTGAAGCATGGCTATTGGAGGTGGAAGAGTGTCTGAACCTTTGAG 729  
Db 64 ACAGCAACCATCAGGCGATTTCCCAATTTGGCAGTGAAGAGGAGTTCAGAAATTTGAAGA 123  
Qy 730 AAAGAGGCTGCTGTTGTAAGGAGAACTTCTGCGAGCAATGTCAGGTATTCCTGGAGA 789  
Db 124 AAAGAACCTGTTGTTCTCAGGAGATGGAGTTGTCCTGTCAGCTATGCTGGAGA 183  
Qy 790 ACTTCTGAATGGAAGAAATGCCAAGTCTCTCTCTCTCTCGAGCAGCAGGATCCCCACTGG 849  
Db 184 ACTACAGAGTGGACTGAGTGCCTGTGGACCTTTGCTCAGTCAGCAGGACAGAGGCGC 243  
Qy 850 CATGTGAGGAGCCCGTGTGGCGGTGGGATCCAGACCGGGAGGTGTACTGTGCCAG 909  
Db 244 GCGAACAGACGCGCCTCTGTGGAGGGGGCATCCAGACCCGAGAGGTGTACTCGGTGCAG 303  
Qy 910 -----AGCGTACCAGCAGCTGCGCGCACTGAGGGGCCAAGGAAGTCTCTAGACCT 957  
Db 304 GCCAACGAAACCTCTCTCACAATTAAGTACCCACAGAAACAAAGAGCTCAAGGCCA 363  
Qy 958 GTGAAAGGCAATTATGTGGGACCGCCCGGTGGCCCTCTCAGCTCTGCAATATCCCT 1017  
Db 364 ATGGACTTAAATATATGCATCGGACCTATCCCTTAATACTACAGCTGTGCCACATTCCT 423  
Qy 1018 TGCTCTACGGACTGCATAGTATCTTCCTGTGTCAGCCTGGGGCTGTGTCATCCATGAAAC 1077  
Db 424 TGTCCTCACTGATGTGAGTTTACCTTGTGTGCTGAGCTTGGGACCTTGTACTTATGAAC 483  
Qy 1078 TGTGTAACCTCAGGGGAAAAAGGATTTAGAACGAGGAGCG-----CCATGTC 1128  
Db 484 TGTAAATGATCAGCAAGGAAAAAGGCTTCAAACTGAGGAAGGGGGGCAATTACCAATGAG 543

Qy 1129 CTCTAGGAATCTACAGGCGCTGCAGGCAATTCCTCATTTTGTGGAGTCTGTTCCTTGT 1188  
Db 544 CCCACTGGAGCTCTGGGGTAACCGGAACCTGCCCTCACTTACTTGAAGCATTCCCTGT 603  
Qy 1189 GAGGATCCAAATGTGTACCGATGGCTGGCAT---CAGAAAGGATCTGTTTCCCTGATCAT 1245  
Db 604 GAAGAGCCTGCCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAAC 663  
Qy 1246 GGAA---AATGTGGCTGGGACATCGTATTCTGAAGGCCGTCTGCCAGATGACCGCGGA 1302  
Db 664 GGAAGGAGTGTGGTCCAGGCCACGCAAGTTCAAGAGGTTGTGTGCATCAACAGTATGGA 723  
Qy 1303 GAAGATGATCAGGAGTCTTTGCCAGTTCCCTCTCTCTCTGAGAGGAGTCTTGTGAA 1362  
Db 724 GAAGAAGTTGACAGACAGCTGTGCAGAGATGCATCTTCCCCTATCCCTGTGGCTGTGAT 783  
Qy 1363 ATTCCCTGCCGAATGACTGTGTGCTGAGCGAGTGGACGGAGTGTCTATCTCTGTTCCAG 1422  
Db 784 GCCCATGCCGGAAGACTGTGTGCTCAGACATGTTCTACGTGTCTCTCTGTCACAC 843  
Qy 1423 TCCTCTTCAAAATAAAACTCAGATGGGAAACAGACACAGGTCAAGAACTATCTCTGC---A 1479  
Db 844 ACCTGCTCAGGGAAAACGACAGAGGGGAAACAGATACGAGCAGCATTCATTCTGSCCTAT 903  
Qy 1480 CTGGCTGGGAGGTGGAAGCCATGTCCCCTTAGTCAAGCTCTCCAGAGCATGCTTTG 1539  
Db 904 GCGGGTGAAGAGGTGGAAATTCGCTGTCCAAATAGCAGTGTCTTTCAGAAAGTACGAAGC 963  
Qy 1540 TGTAAATGACCATTCCTGTATGACGCTTCACCTGGGAGACATCGCTTGGGGCCCTTGTCT 1599  
Db 964 TGTAAATGACATCCTTGCACAGTGTACCACTGGCAACTGTGTCCCTGGGCGCAGTGATT 1023  
Qy 1600 GAGGACATCTGGTAACTGCCCTTAAATGCAACCAATGGCTGGAATGGAAGACCACTGT 1659  
Db 1024 GAGGACACTCAGTATCTGCTTCAACAACTACGACTTGGAAATGGGGAGGCTCTCTGC 1083  
Qy 1660 GGTGTAGCATTACAGCTCGGAGAGTCTTGTGTCAAGAGTCACTGGGCAACAAGTAATG 1719  
Db 1084 TCTGTGCGCATGCAGCAAGAAAGTCACTGTGTGCGAGTCAATGTGGGCCAAGTGGGA 1143  
Qy 1720 ACCAAAGATGTCAGATTTACTGACCTGAAACTGTGCGCCCTCTGTTTCTCCCTATGC 1779  
Db 1144 CCAAAAAAATGTCTGAAAGCCTTCGACCTGAAACTGTGAAGCCTTGTCTGCTCTCTTGT 1203  
Qy 1780 AAAAAGACTGTATTGTCACTGCTTTCAGTCAGTGGGACACCTGCCCAAGGATGCGCAA 1839  
Db 1204 AAGAAGACTGTATTGTGACCCCATATAGTACTGGACATCATGCCCTCTCTGCTGTAA 1263  
Qy 1840 GCAGAAATGCCACAGT---AAAAAGTCTCGATACAGAATCATCTCCAAAGAGCAGCC 1896  
Db 1264 GAAGGGGACTCCAGTATCAGGAAGCAGTCTAGGCATCGGGTCACTTCAGCTGCCAGCC 1323  
Qy 1897 AATGAGGCCAGGAATGCCAGATACCTTATATAGGAGAGAGTGTGAGATGTTTCC 1956  
Db 1324 AACGGGGCCGAGACTGCACAGATCCCCCTCTATGAAGAGAGGCTGTGAGGCACTCCAA 1383  
Qy 1957 TTGTCTCTGTATATCGGTGGAAGCCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAG 2016  
Db 1384 GCGTGCCAAAGCTACAGTGGAGAGACTCACAATGGCGCAGATGCCAATTAGTCTCCTTGG 1443  
Qy 2017 TCTGTCTGGAGGGAATMA CGGGCAGCAGTGAAGCTGTGGAAGGGGTTACAAACAGA 2076  
Db 1444 AGCGTGCAACAGACAGCCCTGGAGCAGAGAGGCTGTGGGCCCTGGGGCAGAGCAAGA 1503  
Qy 2077 GTGTCTCATGATCTCTGTATGACAA CCGGTCAACAGAAATGATGGAATGCTCAAGCAG 2136  
Db 1504 GCCATTACTTGTGCAACAGAGATGGAGCAGGCTGGAATCCATGATGTGCTACAGTAT 1563  
Qy 2137 ACAAAGGCATGCTCTCTTGTGCAAGAAATGCAAGTCCCATCTCGAGAGAGTCAACC 2196  
Db 1564 GCGAGCCCTGTGCCAGCCCTTACCAGGCCCTGCCAGATCCCTGCCAGGATGACTGTCAA 1623



Db 3784 GCTTGCAAAAGCCAAAGAAACCCCAAGAGGCAAAACACACGACTGAACCTTTAACC 3843  
QY 4369 TTAGCCTACGATGGAGACTTAGACATGTAA 4398  
Db 3844 TAGCCTATGATGGAGATGCCGACATGTAA 3873  
RESULT 9  
AX780247  
LOCUS AX780247 5287 bp DNA linear PAT 14-JUL-2003  
DEFINITION Sequence 2404 from Patent WO03039443.  
ACCESSION AX780247  
VERSION AX780247.1 GI:32697241  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1  
AUTHORS Haferlach, T., Schoch, C., Kern, W., Kohlmann, A., Schnittger, S.,  
Dugas, M., Eils, R., Brors, B. and Mergenthaler, S.  
TITLE Novel genetic markers for leukemias  
JOURNAL Patent: WO 03039443-A 2404 15-MAY-2003;  
Deutsches Krebsforschungszentrum (DE);  
Ludwig-Maximilian-Universitaet Muenchen (DE); Haferlach, Torsten,  
PD Dr. Dr. (DE); Schoch, Claudia (DE); Kern, Wolfgang (DE)  
FEATURES  
Location/Qualifiers  
1..5287  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 25.2%; Score 1109.8; DB 6; Length 5287;  
Best Local Similarity 58.4%; Pred. No. 0;  
Matches 2260; Conservative 0; Mismatches 1529; Indels 83; Gaps 16;  
QY 610 TGGAGCCCTGCTCCAGACATGGCTATTGGAGGTGGAAAGGAGTGTCTGCACTTCTTTAG 729  
Db 4 TGGAGCCCTGCTCAAAACATGCCATGACATGTGTGCTCCCTGTCAGGACCTCTGTTAAG 63  
QY 670 AGCGGAACGTGAAGACATGGCTATTGGAGGTGGAAAGGAGTGTCTGCACTTCTTTAG 729  
Db 64 ACAGAACCATCAGCAGTTTCCCATTTGGCAGTGAAGAGGTGTCCAGATTTGAAGAA 123  
QY 730 AAAGAGGCTGATTTGTAAGAGGAATCTTCGAGCAATGTCGCCAGTATCTCTGGAGA 789  
Db 124 AAAGAACCTGTTTGTCTCAAGGAGATGGAAGTGTCTCCCTGTGCCAGCTATGGCTGAGA 183  
QY 790 ACTTCTGAATGAAAGATGCCAGTCTCTCTCCTCTCGAGCAGCAGGATCCCCACTGG 849  
Db 184 ACTACAGATGAGATGAGTGCCTGTGTGACCTTTGTCTAGTCAGCAGGACAGAGCGC 243  
QY 850 CATGTGACGGACCGCTGTGTGGGTGGATCCAGACCCGGGAGGTGTACTGTGCCAG 909  
Db 244 GCAACACAGACGGCCCTCTGTGGAGGGGCATCCAGACCCGAGAGGTGTACTGGCTCAG 303  
QY 910 -----ACGTACAGAGCTGCCGACTGAGGCGCAAGGAGTCTTAGACCT 957  
Db 304 GCCAACGAAACCTCTCTCAATTAAGTACCACAGAAAGAGGCTCAAGCCA 363  
QY 958 GTGGAAGAGGCATTATGTGTGGACCCGCCCTGTGCTCCTCAGCTCTGCAATATCCCT 1017  
Db 364 ATGGACTTAAATATGACCTGGACCTATCCCTAATACTACAGCTGTGCCATTTCT 423  
QY 1018 TGCTACGAGTGCATAGTATCTTCCTGTGTCAGCTGGGCGCTGTGTCATCATGAAAC 1077  
Db 424 TGTCCAACGTGAATGTGAAGTTTCACTTGGTGTGAGCTTTGGGACCTTGTACTTATGAAAC 483  
QY 1078 TGTATGAACCTCAGGGGAAAGGATTTAGACGAGGCGAGCG-----CCATGTC 1128  
Db 484 TGTATGATCAGAGGGAAGAAAGGCTTCAACTGAGGAAGCGGCGCATACCAATGAG 543

QY 1129 CTCATGGAATCTACAGGGCTCGAGGCAATGCCCTCATTTGGTGGAGTCTGTTCTTGT 1188  
Db 544 CCACCTGGAGGCTCTGGGTAAACCGGAAACTGCCCTCACTTACTGGAAGCCATTTCCCTGT 603  
QY 1189 GAGGATCCAATCTGCTACCGATGGCTGGCAT---CAGAAGGAGTCTGTTTCCCTGATCAT 1245  
Db 604 GAAGAGCCTGCCTGTTATGACTGGAAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAAC 663  
QY 1246 GGAA---AATGTGGCTGGGACATCGTATTCTGAAGGCGGTCTGCCAGATGACCGCGGA 1302  
Db 664 GGAAAGGAGTGTGGTCCAGGCGACGAAGTCAAGAGGTGTGTGATCAACAGATGATGGA 723  
QY 1303 GAAGATGTATCAGGGAGTCTTTGCCAGTTCCTCCCTCTCTGAGAGGAAGTCTTTGTGAA 1362  
Db 724 GAAGAAGTTGACAGACAGCTGTGCAGAGATGCCATCTTCCCACATCCCTGTGSCCTGTAT 783  
QY 1363 ATTCCCTGCCGAATGAGCTGTGTGCTGAGCGAGTGGAGCGAGTGTCTATCTCTGTTCCAG 1422  
Db 784 GCCCATGCCCGAAAGACTGTGTGCTCAGCACAATGCTAGCTGTCTCTCTGCTGCACAC 843  
QY 1423 TCTGTTCATAATAAACTCAGATGGGAAACAGACAGGTCAAGAACTATCTCTGCG---A 1479  
Db 844 ACTGCTCAGGGAAGAAACGACAGAGGGAACAGATACGAGCAGATCCATTTCTGSCCTAT 903  
QY 1480 CTGGCTGGGAAGGTGGAAGCCATGTCCCTCTAGTCTAGGCTCTCCAGAGCATCGTTTG 1539  
Db 904 GCGGGTGAAGAGGTGGAATTCGCTGTCCAAATAGCAGTGTCTTGCAGAGAGTACGAAGC 963  
QY 1540 TGTATGACCATTCCTGTATGACGCTTCACTGGGAGACATGCCCTTGGGCGCTTGTCT 1599  
Db 964 TGTATGACCATTCCTTGCACAGTGTACCACTGGCAAACTGGTCTCCCTGGGCGCAGTGCAT 1023  
QY 1600 GAGGACATTTGTTAACTGCCCTTAATGCAACCATTTGGCTGGAATGGAGAACCCAGTGT 1659  
Db 1024 GAGGACACCTCAGTATCGTCTTCAACACAACATACGACTTGAATGGGAGGCTCTCTGC 1083  
QY 1660 GGTGTAGGATTCAGACTCGGAGAGTCTTCTGTGTCAAGAGTCACTGTGGAACAAGTATG 1719  
Db 1084 TCTGTGGCGATCGACAGCAAGAAAGTCACTGTGTGCGAGTCAATGTGGGCGCAAGTNGA 1143  
QY 1720 ACCAAAGATGTCCAGATTTCTACTCGACTGAACTGTGGCGCCCTGTTTCTCCCATGC 1779  
Db 1144 CCCAAATAATGTCTGAAAGCCTTCGACCTGNAACTGTAAAGCCTTGTCTGTTCTTGT 1203  
QY 1780 AAAAAGACTGTATTGTGACTGCTTTAGTGTGAGTGGACACCCCTGCCCAAGAT--GTGCA 1838  
Db 1204 AAGAAGGACTGTATTGNNNNNCATATAGTGACTGACATCATGCCCTCTCTNTGTGTAA 1263  
QY 1839 AGCAGGAATGCCACAGT---AAAACAGTCTCGATACAGAACTCATCCCAAGAGCAGC 1895  
Db 1264 AGAAGGGACTCCAGTATCAGGAAGCAGTGTAGGATCGGGTCACTTCAGCTGCGCAGC 1323  
QY 1896 CAATGGAGGCCAGGAATGCCAGATACCTTATATGAGGAGAGAGTGTGAAGATGTTTC 1955  
Db 1324 CAACGGGGCCGAGACTGCACAGATCCCTCTTATGAGAGAGGCGCTGTGAGGCACCTCA 1383  
QY 1956 CTTGTGTCTGTATTCGGTGGAAAGCAGAGAAATGGAGCCCTTGCATCTTAGTGCAGA 2015  
Db 1384 AGCGTGCCAAAGCTACAGTGGAGACTCACAAATGGCGAGATGCCAATTAGTCCCTTG 1443  
QY 2016 GTCTGTCTGGCAGGGAATAAAGCGGCGAGCAG--TGAAGCCTGTGGAAAGGGGTACAAACA 2074  
Db 1444 GAGCGTGCAACAAGACAGCCCTGAGCAGCAGGAGGCTGTGGGCTGGGCGCAGGCGAA 1503  
QY 2075 GAGCTGTCTCATGCAATCTCTGATGACAACCGGTGAGCAGAAATGATGAAATGCTCTAAGC 2134  
Db 1504 GAGCCATTACTTGTGCGAAGCAAGATGGAGCAGGCTGGAATCCATGAGTGCCTACAGT 1563  
QY 2135 AGACAAAAGGCATGCTCTCTTGTGCAAGAAATGCACAGTCCCATGTGAGAGACTGCA 2194  
Db 1564 ATGAGGCCCTGTGCGAGCCCTTACCGAGGCTGCGAGATCCCTGCGCAGGATGACTGC 1623  
QY 2195 CCTTCACTGCTTGGTCCAAAGTTTACGCCCTCTGCTCCACGAACTGTGAAGCCCAAAAAGTA 2254

Db 1624 AATTGACAGCTGTCACAGTTTCTTCATGCAATGGAGAGCTGTGGTGCAGTTAGGACCA 1683  
QY 2255 GCGCGGACAGCTCACAGGGAAGAGAGGAGAAATCCAGGATTTCTGACCTTT 2314  
Db 1684 GAAAGCCGACTCTTGTTGGAAAAAGTAAAGAGGAAAAATGTAAAAATCCCATTTGT 1743  
QY 2315 ACCCTCTAGTGGAGACAGACTATGTCCTTGTGATGAATTTATATCCCACTTATGGAA 2374  
Db 1744 ATCCCTGATGAGACTAGTATTTGCTTGTGACAAATATAATGACACACCTGTGGGA 1803  
QY 2375 ACTGGTCAGATTCGATCTTCCAGAGGCAAGGAGCCCTCACAGGAGCTCGGGGTAC 2434  
Db 1804 ACTGGTCAGACTGTATTTACCAGAGGAAAAAGTGGAAAGTTGCTGGGAATGAAGTAC 1863  
QY 2435 AAGCAGACACCAAGATGTGGAGAGGCTGCGCTTTCGAGCAGTAGCTGTCTGATA 2494  
Db 1864 AAGAGACATCAAGGAATGCGGACAGGATATCGTTACCAGCAATGGCATGTACGATC 1923  
QY 2495 AAAATGGAAGACCTGTTGACCCCTCTCTGACAGCAGCTCTGGTTACATTCAGAAAAAT 2554  
Db 1924 AAAATGSCAGGCTTGTGAAACATCTAGATGTAAACAGCCATGTTACATTTGAGGAGCCT 1983  
QY 2555 GTGTCACTCCCTGCCATTTGATTCAGTAAAGCAATGGTCTAGTTGGGGTCTTGCA 2614  
Db 1984 GCATCATCCCTGCCCCCTCAGACTGCAAGCTCAGTCAGTGGTCCAACTGCTCGCGCTGCA 2043  
QY 2615 GTTCATCTTGTGAATTTGGAGTGAGAAATTCGATCCAAATGCTAAAGAAAAACCTTACA 2674  
Db 2044 GCAAGTCTCTGGAGTGTGTGAAGTTCGTTCTAAATGCTGCGNGAAANCCATATA 2103  
QY 2675 ATGAGGACGACCATGTCCTCAAACTGGATC-----TCAGAAATCAGGTACATGAGGCAG 2728  
Db 2104 ATGAGGAAGGCTTGGCCCCAACTGACCATGTCAACAGGCAAGGTGTATGAGGTTG 2163  
QY 2729 TCCCATGTTACAGTGTGCAATCAGTATTTCTCTGGTGTGTAGAACCTGCTTCTATGCA 2788  
Db 2164 TCCCATGCCACGTGACTGCAACAGTACTATGGGNCACAGAGCCCTGGAGCATCTGCA 2223  
QY 2789 AAATCAACAAATGAGCTGAGTCCCTGCGCTGTGGAGGAGGAAACAATCTTAGGAAATCA 2848  
Db 2224 AGGTGACCTTTGTGAATATCGCGGAGAACTGTGGAGAGGCGTGCAACCCGAAAGTGA 2283  
QY 2849 GATGTGT---GAATCTGCGGATGTTGAAGGTGGAGAGTGGATAGCAACCTGTGCAAC 2905  
Db 2284 GATGCAATGCAAGATACAGCAGATGGCCCTTCTTAAACATGTAGAGGATTACCTCTNTNACC 2343  
QY 2906 AGGATGAAATTTCCCCAGAAACCCAGTCTCTGTTCTCTTATGTCCTCCAAATGAGTGTCA 2965  
Db 2344 CAGAAGAGATGCCCTGGGCTCTAGAGTGTGCAAAATACCATGCCCTGAGGACTGNTGA 2403  
QY 2966 TGTCTAGTGGGACCTTTGGAGCAAAATGCCACAGTCATGCCATCCCCACAAT---GC 3022  
Db 2404 TATCTGANTGGGTCCATGAGACCCAAATGNTTTTGCTTTGCAATCAAGCAGTTTCCGCG 2463  
QY 3023 AGAGAAAGACTCGCCACCTGCTAAGACCATCACTGACTCAAGGACTTGTGCTGAAGCT 3082  
Db 2464 AAAGTCACTGATCCCATCAGCAACACAGCTGATGAAGAAGATCTTGCCCTAATGCTG 2523  
QY 3083 CACAGTGTGAGCCCTTGCTCTGAAATGAAATTTGCTTCCAGTTTCCAGTACAATCTAAACAG 3142  
Db 2524 TTGAGAAAGAACCTGTAACTTGAACAAAACTGTACCACTATGATTATATGTPAACAG 2583  
QY 3143 AGTGGAGCACAATCCAGCTGAGTGAAGAACGACCCCTGTGGTCAAGGCTCAGGACCCGCC 3202  
Db 2584 ACTGGAGTACATGTCAGCTGAGTGAAGGCAATTTGTGGAAATGGAAATGAAGGCAAGGA 2643  
QY 3203 TGTAAAGCTGTGTGTGAGTGGCAAGCAGCTCAGCATGACCAATGTCAGGAGTGTGAGCAGCAT 3262  
Db 2644 TGTGTGATTTGTTCGAAGTGAAGGCAAGTCAAGTGAACCTGAAATATTTGTGAGGCGCTTG 2703  
QY 3263 ATTTGAGAGAGCCCAAGAAATGAGCATTTCCCTGCTTGTGGAAATGCGTGTGCAACTGTC 3322

Db 2704 GCTTGGAGAGAACTGGCAGATGAACAGCTCCTGCATGGTGGAAATGCCCTGTGAACGTGC 2763  
QY 3323 AGCTCTCAGGCTGGACGCTTTGGACAGAGTGTTCACAGACCTGTGGCCATGGAGGTGAA 3382  
Db 2764 AGCTTCTGATTTGGTCTCCTTTGGTCAGAAATGTTCTCAAACATGTGGCTCACAGAAAA 2823  
QY 3383 TGAGCCGACCTCGATTATCATTTATGCCAACCCAGAGAGAGGAGCGCCATGCCCCACAG 3442  
Db 2824 TGATCCGAAGACGAACAGTGAACCCAGCCCTTTCAAGGTGATGGAAGACCATGSCCTTCCC 2883  
QY 3443 AGCTTACCACGAGAGAAAACTCGCCAGTGACCCCTCTGTACAGCTGGGTCTTGGCAACT 3502  
Db 2884 TGATGGACAGTCCAAACCCCTGCCAGTGAAGCTTGTATCCGGTGGCAATATGGCCAGT 2943  
QY 3503 GGTCTGCACTAAATTTGGAGGTTGAGACTGTGGGAAGAGATTCAGATCCGAGCCTTT 3562  
Db 2944 GGTCTCCATGCCAATGTGAGAGGCCCTGAGTGTGAGAGGAGCCAGAACAAAGGAACATTT 3003  
QY 3563 CTTGCATGCTCCACAGTGGTTCAATATCTCATGAGCTGGACGTGTGAGATGCACGTG 3622  
Db 3004 CTTGTGTAGTAAGTGTGGGTGAGTGAATTTTCAAGAGTGGTGTGATGAGGAATCT 3063  
QY 3623 GTGGAGAAATGCCCTT-----TCAGGACAGCATCCTGAAGCAGCTGTGTT 3667  
Db 3064 GTGCTGACATTTGAATCATTTATAGATGGTAATAAAATATGGTTCTGGAGGAATCTTGCA 3123  
QY 3668 CTGTGCCCTTGGCCAGGAGACTGCCATTTACAGAAATGGTGTGAGTGGAGCAGACATGTGAT 3727  
Db 3124 SCCAGCTTGGCCAGGTGACTGTTATTTGAAGGACTGGTCTTCTCTGGAGCCTGTGTGAGC 3183  
QY 3728 TAACTGCAATTTGATGGAGAAAGCTTTGAGACTGTGGCCGCGCAGTCTAGATCAAGGACTT 3787  
Db 3184 TGACCTGTGTAATGGTGAAGTCTAGGCTTTGGTGAATACAGGTGAGATCCAGACCGG 3243  
QY 3788 TTATAATTCAGTCTTTTGGAGAACCAAGACAGCTGCCCCCAACAGGTTCTAGAAACAGGCC 3847  
Db 3244 TGATTATCAAGAACTAGAGAAATCAGCATCTGTGCCCAGAGCAGATGTTAGAAACAAAT 3303  
QY 3848 CTTGTACAGGAGGCAAAATGTTATCACTACATCGAAAGCAAGTCTTTTGGAAACAATAACG 3907  
Db 3304 CATGTTATGATGGACAGTGTCTATGAATATAAATGATGGCCAGTGTCTTGGAGGGCTCTT 3363  
QY 3908 AACGAACTGTATGTCGACGCTTTCAGATGGCGTTAATGTCAAGGAGGCTGTCCCTCCT 3967  
Db 3364 CCCGAACAGTGTGGTGTCAAAGGTGAGATGTAATAATGTAAACAGGGGGTGTCTGTGTA 3423  
QY 3968 AGGCCGCTCTGCTGCCATTCGGCAGTGCATTCAGCTCCAGCTTCAGAAAPACCTTTCTCTACT 4027  
Db 3424 TGAGCCAGCTGATGCGACAGGCTTGTAAACCCAGTGTGTCAACCCCACTCTGACT 3483  
QY 4028 GTACACAGGTTGAGTCTGTGTTGTGAGAGGCTATACAGAGATAATGAAATCAAAATG 4087  
Db 3484 GTAGCGAGACAAAAACATGCCATTTGTGAAGAGGTTACACTGAAGTCAATGCTCTTCTAACA 3543  
QY 4088 GTTTCCTGGATTACTGCA-----TGAAAGTACAGGCTCAGAGGTACAGAGTAAAA 4132  
Db 3544 GCACCTTGTAGCAATGACACTTATCCCTGTTGGTGTATTACCCACCATGGAGACAAAA 3603  
QY 4133 AAGCTGATGTGAAAAACCTTTCTG-GGAAAAACAGACTGTGAATTCAAAAATACATGAT 4191  
Db 3604 GAGGAGATGTGAAAAACAGTCGGGCTGTACATCAACCCCAACCTCCAGTAAACCCAGCAG 3663  
QY 4192 ATTTTAAAGGA---TGGTCTCTTCAACCACTTGATTCAGATGGCCGAGTAAAAATTTGGG 4249  
Db 3664 GACGGGAAGACCTGTTCTTACAGCCATTTGGGCCAGATGGGAGACTAAAGACCTGGG 3723  
QY 4250 TTTATGGCTTTCAAGTGGCGCTTTCTCATCATGATTTTCTTAATATTTTACTTCTTACC 4309  
Db 3724 TTTAGGTGTAGAGCTTGGGCAATTTGTGTACTCATCTTTTATGTTCTCCATGATTTATC 3783  
QY 4310 TGTGTTGCAAGAGCCA---AAACCATCAAAAGCACACCTCCCCCAACAGAGCCTTGA 4366  
Db 3784 TAGCTTGCAAAAGCCCAAGAAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGG 3843

QY 4367 CCTTACCTACGATCGAGACTAGACATGTAA 4398  
 |||||  
 Db 3844 COTTAGCTATGATGAGATCGCGACATGTAA 3875

## RESULT 10

AX747433

## LOCUS

AX747433

## DEFINITION

Sequence 958 from Patent EPI308459.

## ACCESSION

AX747433

## VERSION

AX747433.1 GI:32131821

## KEYWORDS

Homo sapiens (human)

## SOURCE

Homo sapiens

## ORGANISM

Homo sapiens

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## AUTHORS

Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,

Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,

Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and

Masuho, Y.

## TITLE

Full-length cDNA sequences

## JOURNAL

Patent: EP 1308459-A 958 07-MAY-2003;

Helix Research Institute (JP); Research Association for

Biotechnology (JP)

## FEATURES

source

1..3053

Location/Qualifiers

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

## ORIGIN

Query Match

Best Local Similarity

Matches 1309; Conservative

13.2%; Score 580.6; DB 6; Length 3053;

Pred. No. 3.9e-165;

Mismatches 934; Indels 78; Gaps 9;

QY 10 TGCATTGAGAGCTGAAACCGAACTCTGGTTGCAATGAAATATGCGAACACTTTGCCCTT 69  
 |||||  
 Db 730 TGCATCCGAAGAAGACAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 789  
 |||||  
 QY 70 CAGCTCTCTACAGAACAGCTTGCTCATCTCTGCTCCCGGATTTGTAGTACTCTGAG 129  
 |||||  
 Db 790 AAGCTCTCTGGACAGCTTGCTCATCTCTGCTCCAGCAAGATTTGATCTGTCTGAA 849  
 |||||  
 QY 130 TCTTACCATGTCACACTGTAGCAAGGATGTGGGAAGAAATTCAGCAATAGAACTCGC 189  
 |||||  
 Db 850 TTTCTGCTGCTGCTCGAATGCTCCAAAGCTTGGCGAGCGGCTCCAGCACCGACGGT 909  
 |||||  
 QY 190 GCGTTCATAGCTCCCTCTCTTTGGTGTGCAATGTCGAATCTGACTGAGTCAAGA 249  
 |||||  
 Db 910 CATGTGGTGGCGCTCCCGAGTTCGGAGGCTCTGGCTGTCCAAACCTTGACGGATTCAG 969  
 |||||  
 QY 250 GCTGTGATGCTCCCATTTCTCTGCTCTCTGGGGAAGAGGAATATACATTTAGCCTTAA 309  
 |||||  
 Db 970 GTGTG-----CCAAATCCAGTCCATCGAGGCGGAGGAGCTCAGGTACAGCTGCAT 1020  
 |||||  
 QY 310 GTTGACCATGAGTAAATGCAAGCTGCTCATCTTAAAGAAATTAATCCAGCGGAAGA 369  
 |||||  
 Db 1021 GTGGGCGCTGGAGCACCTGTCTCAATGCCCCACTCCGACAAAGTAAGACAAGGAGA 1080  
 |||||  
 QY 370 ACTGTCTGTGATTTTAA-----CTCTGATTCAAAATGAG 402  
 |||||  
 Db 1081 CCGCGGAAGATAAAGAACGGGAAGAGACCGCAGCAAGAGTAAGGATCCAGAGCC 1140  
 |||||  
 QY 403 CGAGTCACCTTTAAACATCAAGTTTACA-----AAGCACATCATCATTCGAAGTCT 453  
 |||||  
 Db 1141 CCGAGCTCTTTTAAAGAAAGAAAGAAACAGGACAGGACAGACAGAGAGACAAATAT 1200  
 |||||  
 QY 454 TGGGCAATAGACATAGTTTATCAACCCGCGAGGTTTCGTGTACAGAGAGTGTGACAA 513  
 |||||  
 Db 1201 TGGGACATCCAGATTGGATATCAGCCAGAGAGGTTATGTGATTAACAGACGGGAAA 1260  
 |||||

QY 514 AATGCTATGTTAAGCCTTTGCTTCAAGATCTTCCCATTTGATGTTTCACTGCTGCTGATC 573  
 |||||  
 Db 1261 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGAGCTTCCAATGACCTTCCAGCTCTGTGTG 1320  
 |||||  
 QY 574 ATGCCCAAGAGCTGTGAACCTCCAGTGGTCTCTCTGGAGCCCTGTCCAAAGACATGC 633  
 |||||  
 Db 1321 ATCACCAGAGTGCAGAGTTTCCGAGTGGTCAAGATGGAGCCCTGCTCAAAACATGC 1380  
 |||||  
 QY 634 CGTTCAGGAGTCTCTTGCAGGATTTAGGACAGAGCCGGAACGTGAAGCACATGGCT 693  
 |||||  
 Db 1381 CATGACATGGTGTCCCTCGAGGCACTCGTGAAGGACACGAACCATCAGGCGATTCCC 1440  
 |||||  
 QY 694 ATTGGAGGTGGAAGAGGTCTCTGAACTTCTTGAGAAAGAGGCTGCTGTAAGA 753  
 |||||  
 Db 1441 ATTGGCAGTGAAGAGGTGTCCAGATTGGAAGAAAGAACCTGTTGTCTCAAGA 1500  
 |||||  
 QY 754 GAATTTCTCAGCAATGTCCAGGTATTCCTCGAGAACTTCTGAATGGAAGAAATGCCAA 813  
 |||||  
 Db 1501 GATGGAGTTGTCCCTGTGCCACGATGCTGAGAACTTACAGAGTGGACTGAGTGCCT 1560  
 |||||  
 QY 814 GTCTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGACGGGACCGTGTGTGC 873  
 |||||  
 Db 1561 GTGACCCCTTTGCTCAGTCAGCAGGACAAAGAGCGCGCAACACAGAGGCCCTCTGTGA 1620  
 |||||  
 QY 874 GGTGGATCCAGACCCGAGGTGTACTGTGTGCCAG-----AGCGTACAGCA 921  
 |||||  
 Db 1621 GGGGCGCATCCAGACCCGAGGTGTACTGTGTGCCAGCAACCAACCTCTCTCACAA 1680  
 |||||  
 QY 922 GCTGCCGCACTGAGGCCCAAGAGTCTCTAGACCTGTGGAAAGGCATTATGTGTGGA 981  
 |||||  
 Db 1681 TTAAGTACCCCAAGAAACAAAGAGCCTCAAGACCAATGGACTTAAATTTATGCACTGA 1740  
 |||||  
 QY 982 CCGGCCCTTGGCTCTCAGCTCTGCAATATCCCTTGTCTTACGAGCTGCATAGTATCT 1041  
 |||||  
 Db 1741 CCTATCCCTAATACACAGCTGTGCCACATCTCTGTCCAACTGAATGTGAAGTTCA 1800  
 |||||  
 QY 1042 TCTGTGTGACCTGGGCTGTGCAATGAACTGTCTATGAACTTCACTGAACTCAGGGGAAAA 1101  
 |||||  
 Db 1801 CTTGTGTGAGCTTGGGACCTTGTACTTATGAACTGTGTAATGATCAGCAAGGAAAAAA 1860  
 |||||  
 QY 1102 GGAATTTAGAACAGGCGAGCG-----CCATGTCTCATGGAATCTACAGGCGCTGCA 1152  
 |||||  
 Db 1861 GGTCTCAACTGAGNAGCGCGCATACCAATGAGCCACTGGAGGCTCTGGGTGTAAC 1920  
 |||||  
 QY 1153 GGCATTTGCCCTCATTTGGTGGAGTCTGTCTTGTGAGGATPCCAAATGTGTPACCGATGG 1212  
 |||||  
 Db 1921 GGAACCTGCCCTCACTTACTGGAAGCCTTCCCTGTGAAGGCTGCTGTGTTATGACTGG 1980  
 |||||  
 QY 1213 CTGGCA---TCAGAGGAGTCTGTTTCCCTGATCATGGAA---AATGTGGCTGGGACAT 1266  
 |||||  
 Db 1981 AAGCAGTGAGACTGGGAACTGCGAGCCAGATAACGGAAGAGTGTGGTCCAGGACG 2040  
 |||||  
 QY 1267 CGTATTTCAAGGCGCTGTGCCAGAAATGACCGGAGAGAGATGTATCAGGGAGTCTTTTC 1326  
 |||||  
 Db 2041 CAAATTTCAAGAGGTTGTGTGCATCAACAGTGTGAGAGAAAGTTGACAGACAGCTGTGC 2100  
 |||||  
 QY 1327 CAGATTTCCCTCTCTCTGAGAGGAGTCTTGTGAAATTCCTGCGGATGAGACTGTGTG 1386  
 |||||  
 Db 2101 AGAGATGCCATTTCCCATCCCTGTGGCTGTGATGCCCCCATGCCGAAAGACTGTGTG 2160  
 |||||  
 QY 1387 CTGAGCGAGTGAACGAGGTGTCATCTGTTCCAGTCTCTGTTCAATTAATAAACTCAGAT 1446  
 |||||  
 Db 2161 CTCAGACATGTTCTACGTGCTCTCTCTGCTCACACACTGCTCAGGGAAGACGACAGAA 2220  
 |||||  
 QY 1447 GGGAAACAGACCGAGTCAAGAACTATCTGTGGCACTGGCTGGG---GAGGTGGAAGCCA 1503  
 |||||  
 Db 2221 GGGAAACAGATACGACGACGATCCATTTCTGGCTATGCGGGTGAAGAGGTTGAAATTCG 2280  
 |||||  
 QY 1504 TGTCCCTCTAGTCAGCTCTCCAAGAGCATCTTGTGTGTAATGACCATTCCTGTGTGACG 1563  
 |||||  
 Db 2281 TGTCCAAATAGCAGTGTCTTGTGAAGAGTACGAGCTGTATGAGCATCTCTTGTGACAGT 2340  
 |||||  
 QY 1564 CTTACTGGGAGACATCGCCTTTGGGGCCCTTTGTTCTGAGGACACATTTGGTAACTGCCCTT 1623  
 |||||

Db 2341 TACACATGGCAAACTGGTCCCTGGGGCACTGATGAGACACCTCAGTATCGTCTTC 2400  
 QY 1624 AATGCAACCACTGGCTGGAATGAGAAAGCACCTGTGGTGTAGGCATTCAGACTCGGAGA 1683  
 Db 2401 AACACAACACTAGCACTGGAAATGGGAGGCTCTGTCTGTGCGCATGAGACAAGAAA 2460  
 2Y 1684 GTCTTCTGTCTCAAGAGTCACTGGGCAAGTAATGACCAAAAGATGTCCAGATTTACT 1743  
 Db 2461 GTCATCTGTGTGGAGTCAATGTGGGCAAGTGGGACCCCAAAAATGTCTGAAAGCCTT 2520  
 2Y 1744 CGACCTGAACACTGTGGCCCTCTTTCTCCCATGCAAAAAGCACTGTATGTGACTGCT 1803  
 Db 2521 CGACCTGAACACTGTAGGCTTGTCTCTTCTGTGAAGAGCACTGTATGTGACCCCA 2580  
 2Y 1804 TTCACTGAGTGGACACCTCCCAAGAGATGTGCCAAGCAGGAATCCACAGT---AAAA 1860  
 Db 2581 TATAGTGACTGGACATCATGCCCTCTTCTGTGAAGAGAGGACTCCAGTATCAGGAAG 2640  
 2Y 1861 CAGTCTCGATACAGATCATCATCCAGAGCAGCAATGAGGCCAGGATGCCAGAT 1920  
 Db 2641 CAGTCTAGGATCGGTCATCATTCAGCTCCAGCAACCGGGGCGAGACTGCACAGAT 2700  
 QY 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCCTTGTGTCTCTATATCGTGGAAAG 1980  
 Db 2701 CCCCTCTATGAAGAGAGGCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGTGGAG 2760  
 QY 1981 CCACAGAAATGGAGCCCTTCATCTTAGTCCAGAGTCTGTCTGGCAGGAATPACGGGC 2040  
 Db 2761 ACTCACAAATGGCGCAGATCCCAATTAGTCCCTTGGAGCGGTGCAACAAGACAGCCCTGGA 2820  
 QY 2041 AGCAGTGAAGCTGTGGAAGAGGCTTACAAACAGAGCTGTCTCATGCTCTCTGATGAC 2100  
 Db 2821 GCACAGAAAGCTGTGGCCCTGGCGACAGCAGAGCCATTAATCTGTGCGAAGCAAGAT 2880  
 2Y 2101 AACCGTTCAGAGAAATGATGAATGGAATGCTCAAGCAGACAAACGGCATGCTCTCTTGTG 2160  
 Db 2881 GGAGGACAGGCTGGAATPCCATGATGCTTACAGTATGACAGGCTGTGCGAGCCCTTACC 2940  
 2Y 2161 CAAGAATGACAGTCCCATGTCGAGAGAGCTGCACCTTCACTGCTGTGGTCCAGTTTACG 2220  
 Db 2941 CAGGCTCCAGATCCCTGCCAGATGATGATCAATGACAGCTGGTCCCAAGTTTCT 3000  
 2Y 2221 CCTGTCTCCAGACTGTGAAGCCACAAAAGTAGGGCGG 2261  
 Db 3001 TCATGCAATGGAGACTGTGTGTCAGTTAGCAGCAAGAGCG 3041

RESULT 11  
 AK092252  
 LOCUS Homo sapiens cDNA FLJ34933 fis, clone NT2RP7005669. PRI 15-JUL-2002  
 DEFINITION  
 ACCESSION AK092252  
 VERSION AK092252.1 GI:21750796  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
 AUTHORS Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S., Ono, Y., Hotta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A., Oshina, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Negahara, K., Masuho, Y., Nagai, K. and Isogai, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 2 (bases 1 to 3053)  
 REFERENCE  
 AUTHORS Isogai, T. and Yamamoto, J.

TITLE JOURNAL  
 COMMENT  
 FEATURES source  
 1.3053  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP7005669"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP7"  
 /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal precursor cells after 5-weeks retinoic acid (RA) induction."  
 ORIGIN  
 Query Match 13.2%; Score 580.6; DB 9; Length 3053;  
 Best Local Similarity 56.4%; Pred. No. 3.9e-165;  
 Matches 1309; Conservative 0; Mismatches 934; Indels 78; Gaps 9;  
 QY 10 TGCATTCAGAAGCTGAACCGCACTGTGGTTCGAATGAAATATCGGAACACTTTGCCCTT 69  
 Db 730 TGCATTCAGAAGCAACAGACATTCCTGGGAGATATCATCTGTAGTACTTTTGAGCCC 789  
 QY 70 CAGCCTCTCAAGAACAGCTTGGCTCATTCCTTGTCCCGGATTTGTAGTATCTGAG 129  
 Db 730 AAGCCTCTCTCGAGCAGCTTGGCTCATTCCTTGCAGCAAGATTGCATCTGTCTGAA 849  
 QY 130 TCTTACCATGCTCAACTGAGCAAGGATGTGGAGAAATTCAGCATAGAACTGCG 189  
 Db 850 TTTCTGCTGCTCGAATGCTCCAGACCTCGCGCAGCGGCTCCAGACCGGAGCGGT 909  
 QY 190 CGGTCATAGTCTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGAGTCAAGA 249  
 Db 910 CATGTGTGGCGCTCCGAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGAGTTCAG 969  
 QY 250 GCCTGTGATGCTCCCATTTCTGTCTCTTGGGAAGAGGATATACATTTAGCTTTAAG 309  
 Db 970 GTGTG-----CCAAATCCAGTCCATGCGAGCGCGAGGAGCTCAGGTACAGCCTGCAT 1020  
 QY 310 GTTGACCATGGAGTAATGCAGACTGCTCATCTTAAAGAAATTAATCCAAAGCGGAGA 369  
 Db 1021 GTGGGCTCTGGAGCACTGCTCAATGCCCACTCCCGACAGTAAGACAAGCAAGAGA 1080  
 QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAAATGAG 402  
 Db 1081 CGCGGAAAGATAAAGAACCGGAAAGACCGCAGCAAGAGGATTAAGATCCAGAAGCC 1140  
 QY 403 CGAGTCACCTTTAAACATCAAGTTAGA-----AAGCACATCATCTTCGAAGTCT 453  
 Db 1141 CCGGAGCTTATTAGAAAAGAGAAACAGAAACAGGAGCAAGACAGACAGAAACAATAT 1200  
 QY 454 TGGGCAATAGAGATAGGTTATCAAAACCGGAGGTTTGTGTACAAAGAGTATGGACAA 513  
 Db 1201 TGGGACATCCAGATTGGATATCAGACCAGAGGTTATGTGTCATTAAACAAGCGGGAAA 1260  
 QY 514 AATGCTATGTTAAGCCTTTGGCTTCAGATTCCTCCCATGACTGTTTCAGTCTGCATC 573  
 Db 1261 GCTGCTGATTTAAGCTTTTGCAGCAAGAGAGGTTCCAATGACCTTCCAGTCTGTGTG 1320  
 QY 574 ATGCCCAAGAGCTGTGAAACCTCCAGTGGTCCCTCTGGAGCCCTGTCTCAAGACATGC 633  
 Db 1321 ATCCAAAGAGTGCAGGTTTCCAGTGGTGCAGTGGAGCCCTGCTCAAAACATGC 1380

Direct Submission  
 Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.  
 Location/Qualifiers  
 1.3053  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NT2RP7005669"  
 /cell\_line="NT2"  
 /cell\_type="teratocarcinoma"  
 /clone\_lib="NT2RP7"  
 /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal  
 precursor cells after 5-weeks retinoic acid (RA)  
 induction."

1684	QY	GTCTTCTGTGTCAAGAGTACAGTGGGAGCAAGTAATACCAAAAGATGTCCAGATTCTACT	1743
2461	Db	GTCAATCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCCAAAAATGTCTGAAAGCCCT	2520
1744	QY	CGACCTGAAACTGTGCGCCCTCTTTTCTCCCATGCCAAAAAGACTGTATTGTGACTGCT	1803
2521	Db	CGACCTGAAACTGTAAAGGCTTGTCTGCTTCTCTTGTAAAGAGGACTGTATTGTGACCCCA	2580
1804	QY	TTCAAGTGTGAGACACCCCTGCCCAAGGATGTGCCAAGCAGGAAATGCCACAGT---	1860
2581	Db	TATAGTAGTGTGACATCATATGCCCTCTTCTGTTAAAGAGGGGACTCCAGTATCAGGAAG	2640
1861	QY	CAGTCTCGATACAGAATCATCATCCAAAGACGAGCCAAATGGAGGCCAGGAAATGCCCAGAT	1920
2641	Db	CAGTCTAGGCATCGGCTCATCATCTAGCTGCCAGCCAAACGGGGCCGAGACTGCCAGAT	2700
1921	QY	ACCTTATATAGGAGAGAGAGTGTGAAAGATGTTTCTTGTGTCTGTATATCGTGGGAAG	1980
2701	Db	CCCTCTATTAAGAGAAAGCCCTGTGAGGCACCTCAAGCGTGCACAAAGCTACAGTGGGAAG	2760
1981	QY	CCACAGAAATGGAGCCCTTTCATCTTGTAGTCCAGAGTCTGTCTGGCAGGGAATAACGGGC	2040
2761	Db	ACTCACAAATGGCGCAGATGCCAATTAGTTCCTTGGAGCGTGCACACAGACGCCCTGGA	2820
2041	QY	AGCAGTGAAGCCCTGTGGAAAGGGGTTTACAAACAGAGAGCTGTCTCATGTCATCTCTGATGAC	2100
2821	Db	GCACAGAAAGGCTGTGGGCTCGGCGACAGGCAAGAGCCATTACTTTGTCCGAAGCAAGAT	2880
2101	QY	AACCGGTTCAGCAGAATGATGAAATGCCCTCAGACGACAGCAAAACGGCATGCTCTCTTTGTG	2160
2881	Db	GGAGGACAGCGTGGAAATCATCAGTGCCTTACAGTATGCAGGCGCCTGTGCGAGCCCTTACC	2940
2161	QY	CAAGAATGCACAGTCCCATGTCTCGAAGAGACTGCACCTTCACCTGTGGTCCAAAGTTTACG	2220
2941	Db	CAGGCGTCCAGATCCCTGCCAGATGACTGTCAATTGACACAGCTGTCTCAAGTTTTCT	3000
2221	QY	CCCTGTCTCAGAACTGTGAAGCCAAAAAGTAGCGCGCG	2261
3001	Db	TCAATCAATGGAGACTGTGTGTCAGTTAGTACCAGAAAGCG	3041

RESULT 12	AC011231/c
LOCUS	AC011231
DEFINITION	Homo sapiens BAC clone RP11-13K3 from 2, complete sequence.
ACCESSION	AC011231
VERSION	AC011231.7
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	AC011231 119205 bp DNA linear PRI 07-NOV-2001
LOCUS	Homo sapiens BAC clone RP11-13K3 from 2, complete sequence.
DEFINITION	Homo sapiens BAC clone RP11-13K3 from 2, complete sequence.
ACCESSION	AC011231
VERSION	AC011231.7
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 119205)
TITLE	Sulston,J.B. and Waterston,R.
JOURNAL	Toward a complete human genome sequence
MEDLINE	Genome Res. 8 (11), 1097-1108 (1998)
REFERENCE	99083792
AUTHORS	9847074
TITLE	2 (bases 1 to 119205)
JOURNAL	Harris,A. and Maupin,R.
MEDLINE	The sequence of Homo sapiens BAC clone RP11-13K3
REFERENCE	Unpublished
AUTHORS	3 (bases 1 to 119205)
TITLE	Waterston,R.H.
JOURNAL	Direct Submission
MEDLINE	Submitted (04-OCT-1999) Genome Sequencing Center, Washington
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
AUTHORS	4 (bases 1 to 119205)
TITLE	Waterston,R.H.
JOURNAL	Direct Submission
MEDLINE	Submitted (23-MAY-2001) Genome Sequencing Center, Washington
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA





KEYWORDS	TITLE
SOURCE	JOURNAL
ORGANISM	COMMENT
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo.

1 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone RP11-41A18

2 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Chesep, Y., Colangelo, M., Collins, S., Callmors, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, F., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lechoczky, J., Levine, R., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melnick, J., Meneus, D., Minova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schaefer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 7, 2000 this sequence version replaced gi:7342347.

All repeats were identified using RepeatMasker:

1 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone RP11-41A18

2 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Chesep, Y., Colangelo, M., Collins, S., Callmors, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, F., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lechoczky, J., Levine, R., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melnick, J., Meneus, D., Minova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schaefer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 7, 2000 this sequence version replaced gi:7342347.

All repeats were identified using RepeatMasker:

1 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone RP11-41A18

2 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Chesep, Y., Colangelo, M., Collins, S., Callmors, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, F., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lechoczky, J., Levine, R., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melnick, J., Meneus, D., Minova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schaefer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 7, 2000 this sequence version replaced gi:7342347.

All repeats were identified using RepeatMasker:

1 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone RP11-41A18

2 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Chesep, Y., Colangelo, M., Collins, S., Callmors, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, F., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lechoczky, J., Levine, R., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melnick, J., Meneus, D., Minova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schaefer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 7, 2000 this sequence version replaced gi:7342347.

All repeats were identified using RepeatMasker:

1 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone RP11-41A18

2 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Chesep, Y., Colangelo, M., Collins, S., Callmors, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, F., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lechoczky, J., Levine, R., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melnick, J., Meneus, D., Minova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schaefer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jun 7, 2000 this sequence version replaced gi:7342347.

All repeats were identified using RepeatMasker:

1 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 2, clone RP11-41A18

2 (bases 1 to 167462)

Barren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Chesep, Y., Colangelo, M., Collins, S., Callmors, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, F., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lechoczky, J., Levine, R., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Melnick, J., Meneus, D., Minova, T., Miranda, C., Mienga, V., Morrow, J

```
* 8018 8117: Gap of 100 bp
* 8118 10466: contig of 2349 bp in length
* 10466: Gap of 100 bp
* 13763 13763: contig of 3197 bp in length
* 13764 13863: Gap of 100 bp
* 13864 15896: contig of 2033 bp in length
* 15897 15996: Gap of 100 bp
* 15997 19987: contig of 3991 bp in length
* 19988 20087: Gap of 100 bp
* 20088 24063: contig of 3976 bp in length
* 24064 24163: Gap of 100 bp
* 24164 26824: contig of 2661 bp in length
* 26825 26924: Gap of 100 bp
* 26925 30928: contig of 4004 bp in length
* 30929 31028: Gap of 100 bp
* 31029 34504: contig of 3476 bp in length
* 34505 34604: Gap of 100 bp
* 34605 40409: contig of 5805 bp in length
* 40410 40509: Gap of 100 bp
* 40510 46642: contig of 6133 bp in length
* 46643 46742: Gap of 100 bp
* 46743 55317: contig of 8575 bp in length
* 55318 55417: Gap of 100 bp
* 55418 64038: contig of 8621 bp in length
* 64039 64138: Gap of 100 bp
* 64139 71781: contig of 7643 bp in length
* 71782 71881: Gap of 100 bp
* 71882 81417: contig of 9536 bp in length
* 81418 81517: Gap of 100 bp
* 81518 102057: contig of 20540 bp in length
* 102058 102157: Gap of 100 bp
* 102158 121599: contig of 19442 bp in length
* 121600 121699: Gap of 100 bp
* 121700 142730: contig of 21030 bp in length
* 142730 142829: Gap of 100 bp
* 142830 167462: contig of 24633 bp in length.
* 142830 Location/Qualifiers
* 1. .167462
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="2"
  /map="2"
  /clone_lib="RP11-41418"
  /clone_lib="RP11-11 Human Male BAC"
  /notes="assembly_fragment"
  1. .1293
  /notes="assembly_fragment"
  1394. .3228
  /notes="assembly_fragment"
  3329. .5140
  /notes="assembly_fragment"
  5241. .8017
  /notes="assembly_fragment"
  8118. .10466
  /notes="assembly_fragment"
  10567. .13763
  /notes="assembly_fragment"
  13864. .15896
  /notes="assembly_fragment"
  clone_end:77
  vector_side:left"
  15997. .19987
  /notes="assembly_fragment"
  20088. .24063
  /notes="assembly_fragment"
  24164. .26824
  /notes="assembly_fragment"
  clone_end:SP6
  vector_side:right"
  26925. .30928
  /notes="assembly_fragment"
  31029. .34504
  /notes="assembly_fragment"
```

FEATURES

```
misc_feature 34605. .40409
  /note="assembly_fragment"
misc_feature 40510. .46642
  /note="assembly_fragment"
misc_feature 46743. .55317
  /note="assembly_fragment"
misc_feature 55418. .64038
  /note="assembly_fragment"
misc_feature 64139. .71781
  /note="assembly_fragment"
misc_feature 71882. .81417
  /note="assembly_fragment"
misc_feature 81518. .102057
  /note="assembly_fragment"
misc_feature 102158. .121599
  /note="assembly_fragment"
misc_feature 121700. .142729
  /note="assembly_fragment"
misc_feature 142830. .167462
  /note="assembly_fragment"
ORIGIN
Query Match 12.0%; Score 527; DB 2; Length 167462;
Best Local Similarity 100.0%; Pred. No. 1.8e-148;
Matches 527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTGCGCTGCATTCAGAGCTGACCGAACTGCTGCTTGCATAATGAAATATGCGAACAC 60
DB 145114 ATGTGCGCTGCATTCAGAGCTGACCGAACTGCTGCTTGCATAATGAAATATGCGAACAC 145055
QY 61 TTGGCCCTTCAGCTCTCTACAGACAGCTTGCCTCATTCCTTCTCCCGGATTTGTA 120
DB 145054 TTGGCCCTTCAGCTCTCTACAGACAGCTTGCCTCATTCCTTCTCCCGGATTTGTA 144995
QY 121 GTATCTGAGTTCTTACCATTGCTTCAACTGTAGCAAGGATGTGGGAAGAAATTCACGACAT 180
DB 144994 GTATCTGAGTTCTTACCATTGCTTCAACTGTAGCAAGGATGTGGGAAGAAATTCACGACAT 144935
QY 181 AGAACTCGCGGTCATAGCTCCCTCTCTTTGGTGGTTGCAATCTCCAAATCTGACT 240
DB 144934 AGAACTCGCGGTCATAGCTCCCTCTCTTTGGTGGTTGCAATCTCCAAATCTGACT 144875
QY 241 GAGTCAAGAGCTGTGATGCTCCCATTTCTCTGCTCTTGGGGAAGAGGAATATACATTT 300
DB 144874 GAGTCAAGAGCTGTGATGCTCCCATTTCTCTGCTCTTGGGGAAGAGGAATATACATTT 144815
QY 301 AGCCTTAAGTTGGACCATGGAGTAATGCGAGTGCCTCATCTTAAAGAAATTAATCCA 360
DB 144814 AGCCTTAAGTTGGACCATGGAGTAATGCGAGTGCCTCATCTTAAAGAAATTAATCCA 144755
QY 361 AGCGGAAGAACTGTTCTGGATTTTAACTCTGATTCAATGAGCGAGTCACTTTAAACAT 420
DB 144754 AGCGGAAGAACTGTTCTGGATTTTAACTCTGATTCAATGAGCGAGTCACTTTAAACAT 144695
QY 421 CAAAGTTACAAAGACATCATCATTCGAAAGTCTTGGGCAATAGAGATAGGTATCAAAACC 480
DB 144694 CAAAGTTACAAAGACATCATCATTCGAAAGTCTTGGGCAATAGAGATAGGTATCAAAACC 144635
QY 481 CGCGAGGTTCTGTGTACAGAGTGTGCAAAATGCTATGTTAAG 527
DB 144634 CGCGAGGTTCTGTGTACAGAGTGTGCAAAATGCTATGTTAAG 144589
RESULT 14
AC111611
LOCUS AC111611.4 Gi:24942115
DEFINITION Rattus norvegicus clone CH230-237P17, *** SEQUENCING IN PROGRESS
ACCESSION AC111611
VERSION AC111611.4
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

# REFERENCE AUTHORS

1 (bases 1 to 224789)  
Murny D, Marie, Metaker, M, Lee, Abranzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, B., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregoriss, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Guevarne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, I., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mapua, B., Mapua, K., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelme, O., Okwou, G., Olarunpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldinger, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steime, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

# TITLE JOURNAL

2 (bases 1 to 224789)  
Worley, K.C.  
Direct Submission  
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 224789)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

# REFERENCE AUTHORS

On Nov 13, 2002 this sequence version replaced gi:22857292.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: CH230-237P17  
Center clone name: CH230-237P17  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 209106 bases at least Q40  
Consensus quality: 213906 bases at least Q30  
Consensus quality: 216832 bases at least Q20  
Estimated insert size: 218750; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 224789: contig of 224789 bp in length.  
Location/Qualifiers  
1..224789  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-237P17"

# FEATURES Source

Query Match 8.2%; Score 362.4; DB 2; Length 224789;  
Best Local Similarity 80.7%; Pred. No. 4.4e-98;  
Matches 423; Conservative 0; Mismatches 101; Indels 0; Gaps 0;  
QY 4 GTGCGCTGCAATTCAGAGCTGACCGAATGTTGTTGCAATGAAATATGCAACACTTT 63  
Db 43659 GTGCGCTGCTTTCAAGAGCTCAACAGAACCAATGTTGCGCCAGTGTGAACACTTT 43718  
QY 64 GCCCTTCAGCTTCTACAGAACAGAGCTTGCCTCATCTTGTCCCGGATTTGTAGTA 123  
Db 43719 GTTCTCAGCTTCCACAGAGCAAGCTTGCCTCATCTTGTCCCGGATTTGTAGTA 43778  
QY 124 TTGAGTTCTTACCATGTTCCAACTGTAGCAAGGATGTGGAGAAATTTGAGCATAGA 183  
Db 43779 TCTGAGTTCTCCCGGTGTCCACCTGCTTGAAGGATGTGGAGAAATTTGAGCATAGA 43838  
QY 184 ACTCGCGCGGTTCATAGCTCCCTCTCTTTTGTGTGTTTGCATGTCCTCAATCTGAG 243  
Db 43839 ACTCGGTGCGCATCGACCCCTCTGTATGAGGTCTGCACTGCCCAATCTCACTGAG 43898  
QY 244 TCAAGAGCTGTGATGCTGCCCATTTCTCTGTCTCTTGGGAGAGGAATATACATTTAGC 303  
Db 43899 TCCAGAGCTGTGAGGCTCCCAATTTCTCTGTCTCTTGGGAGAGGAATATTTCTTCAGC 43958  
QY 304 CTTAAGTTTGACCAATGGAGTAAATGCAAGCTGCTCATCTTAAAGAAATTAATCCAGC 363  
Db 43959 CTTAAGTTGGGACCAATGGAGCAAGTGTAGCTTCCCTCACCTTAAAGAGCCGAGCTCGC 44018  
QY 364 GGAAGAACTGTTCTGGATTTTAACTCTCTGATTCAAATGAGCGAGTCACCTTTAAACATCAA 423

# ORIGIN

Db 44019 GGAAGAAATGTCAGGATTTAGCTCTGATTCAACAGCAGCAAGTCACCCCTAACACATCAG 44078

QY 424 AGTTACAAAGCACATCATCTTCGAGCTTTGGCAATAGAGATAGATTATCAACCCCG 483

Db 44079 AGTTACAAAGCACAGCACCTCCAGCTTGGGCTATAGAGATAGATTTCACAAACAGG 44138

QY 484 CAGGTTCTGTTACAGAAAGTGTGGACAAAAATGCTATGTTAAAG 527

Db 44139 CAGGTGTGGTGCACAGAGTGTGGAGAAAGCCCTGTGTTAAAG 44182

RESULT 15

AC123741/c

LOCUS AC123741.4 GI:30017756

DEFINITION HTG; HTGS PHASE1; HTGS DRAFT.

VERSION HTG; HTGS PHASE1; HTGS DRAFT.

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 259236)

AUTHORS Birren, B., Nusbaum, C. and Lander, E.

TITLE Mus musculus, clone RP24-113A5

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 259236)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 259236)

AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Ruchpaka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,

TITLE  
JOURNAL

## COMMENT

Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 17, 2003 this sequence version replaced gi:28882446.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
-----  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBX  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
-----  
Project Information  
Center project name: L26529  
Center clone name: 113\_A\_5  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 655: contig of 655 bp in length  
\* 1 656 755: gap of 100 bp  
\* 756 1434: contig of 679 bp in length  
\* 1435 1534: gap of 100 bp  
\* 1535 7687: contig of 6153 bp in length  
\* 7688 7787: gap of 100 bp  
\* 7788 83305: contig of 75518 bp in length  
\* 83306 83406: gap of 100 bp  
\* 83406 259236: contig of 175831 bp in length.

FEATURES  
source

1. 259236  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone\_lib="RP24-113A5"  
/clone\_lib="RPCI-24 Male Mouse BAC"

## misc\_feature

1. 655  
/note="assembly\_fragment"

## misc\_feature

756..1434  
/note="assembly\_fragment"

## misc\_feature

1535..7687  
/note="assembly\_fragment"

## misc\_feature

7788..83305  
/note="assembly\_fragment"

## misc\_feature

83406..259236  
/note="assembly\_fragment"

clone\_end:T7  
vector\_side:right

## ORIGIN

Query Match 8.2%; Score 360.8; DB 2; Length 259236;  
Best Local Similarity 80.5%; Pred. No. 1.4e-97;  
Matches 422; Conservative 0; Mismatches 102; Indels 0; Gaps 0;  
QY 4 GTGGCTGCATTCAGAGCTGCAACGAACTGTGTTGCAATAAATGGAACACTTT 63  
Db 218506 GTGGCTGTCTTTCAGAGCTCAATAGAACCATGGTGTCCATGAGATCTGTGAACACTTT 218447  
QY 64 GCCCTTACGCTCTCTACAGAACAGCTTGCTCTATTCCTTGTCCTCCGGGATTGTGTAGTA 123  
Db 218446 GCTCTTAGCCCCCAGAGCAAGCCTGCTCTATTCTTCTTCCCCCAGGAGCTGCTGTGTA 218387  
QY 124 TCTGAGTTCTTACCATGCTCAACTGTAGCAAGGATGTGGAGAAATTCAGCATAGA 183  
Db 218386 TCTGAGTTCTTCCCCATGCTCCACCTGCCCCGAGAGTGTGGAGAAATTCAGCATAGA 218327  
QY 184 ACTGCGCGGTCTAGCTCCCCCTCTCTTTGGTGGTTGCAATGTCCAATCTGACTGAG 243

Db 218326 ACTCGTGTGGCCATCGGCCCCCTCTGTATGGAGGTCTGAGTGTCAAATCTCACGGAG 218267  
Qy 244 TCAAGAGCCTGTGATGTCCCAATTCCTGTCTCTTGGGAAGAGGAATATACATTTAGC 303  
Db 218266 TCCAGAGCCTCGAGGCTCCAGTTTCTGTCTCTTGGGAAGAGGAATATTCCTTCAGC 218207  
Qy 304 CTTAAGGTTGGACCATGGAGTAAATGCAGACTGCTCATCTTAAAGNAATTAATCCAAGC 363  
Db 218206 CTAAGGTTGGACCATGGAGCAATGTAGACTCCCTCACCTTAAGGAAGTCGACCTCAGC 218147  
Qy 364 GGAAGAACTGTCTGGATTTTAACTCTGATTCAAATGAGCGAGTCACCTTTAAACATCAA 423  
Db 218146 GGAAGAAATATCCAGGATTTTAGCTCTGATTCRAACGAGCAAGTCACCTTAAACATCAG 218087  
Qy 424 AGTTACAAGCACATCATTCGAAGTCTTGGCAATAGAGTAGGTTATCAAAACCCGG 483  
Db 218086 AGTTACAAGCACACCCACCACTCCAGCCTGGGGATGTAGTAGGGTTTCAAAACCCAGG 218027  
Qy 484 CAGGTTTCGTGTACAAGAAGTGTGACAAAAATGCTATGTTAAG 527  
Db 218026 CAGGTGTGGTGTACAGAGTGTGGAAGAAACGCCCTGTTAAG 217983

Search completed: February 23, 2004, 17:26:01  
Job time : 16550.2 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 10:46:45 ; Search time 1531.36 Seconds  
(without alignments)  
12200.602 Million cell updates/sec

Title: US-10-022-710-1

Perfect score: 4398

Sequence: 1 agggcgctgattcagaa.....atggagacttagacatgtaa 4398

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_29Jan04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002s:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	4380.4	99.6	5847	9 ADE07142	Novel cod
2	2235.8	50.8	3382	6 AAL43703	Human lar
3	1232.8	28.0	4810	7 ACC72174	Human NOV
4	1232.8	28.0	6317	6 AAD38694	Human LP2
5	1232.8	28.0	6373	4 AAF30188	Clone 214
6	1232.8	28.0	6373	7 ACC72162	Human NOV
7	1232.8	28.0	6373	7 ACC72171	Human NOV
8	1232.8	28.0	6373	7 ACC72168	Human NOV
9	1229.6	28.0	6378	4 AAI58451	Human pol
10	1229.6	28.0	6378	8 ADB48422	Novel hum
11	1190.4	27.1	6292	4 AAI60237	Human pol
12	1177.4	26.8	4647	7 ACC72167	Human NOV
13	1177.4	26.8	4881	7 ACC72165	Human NOV
14	1133.4	25.8	4879	7 ACC72166	Human NOV
15	580.6	13.2	3053	9 ADB62804	Human CDN
16	541.4	12.3	1713	7 ACC72164	Human NOV
17	540.2	12.3	1732	7 ACC72172	Human NOV
18	512.4	11.7	516	6 ABK70149	Human lun
19	505.8	11.5	1228	5 AAS69250	DNA encod
20	460	10.5	744	4 AAK43667	Human bon
21	460	10.5	744	4 AAK17814	Human bra
22	436.6	9.9	456	5 AAS69249	DNA encod
23	402.8	9.2	1401	7 ACC72163	Human NOV

## ALIGNMENTS

### RESULT 1

ADAE07142  
ID ADE07142 standard; DNA; 5847 BP.

XX AC ADE07142;

XX DT 29-JAN-2004 (first entry)

XX DE Novel coding sequence (useful for identifying genetic disorders) #208.

XX KW novel gene; novel protein; tissue marker; molecular weight marker;

XX KW chromosome marker; genetic disorder; gene; ds.

XX OS Unidentified.

XX PN WO2003054152-A2.

XX PD 03-JUL-2003.

XX PP 10-DEC-2002; 2002WO-US039555.

XX PR 10-DEC-2001; 2001US-0339739P.

XX PR 11-DEC-2001; 2001US-0339453P.

XX PR 14-MAR-2002; 2002US-0365091P.

XX PR 12-APR-2002; 2002US-0372381P.

XX PR 12-APR-2002; 2002US-0372615P.

XX PR 22-APR-2002; 2002US-00128558.

XX PR 24-APR-2002; 2002US-0376045P.

(HYSE-) HYSRQ INC.

XX PA Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

XX PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;

XX PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX DR WPI; 2003-569235/53.

XX DR P-PSDB; ADE08053.

XX PT New polynucleotides, useful for expressing recombinant proteins for

XX PT analysis, characterization or therapeutic use, or as markers for tissues

XX PT in which the corresponding protein is preferentially expressed.

XX PS Claim 1; SEQ ID NO 208; 1177pp; English.

XX XX

XX CC The invention comprises the amino acid and coding sequences of novel

Acc72170 Human NOV  
Acc72169 Human NOV  
Acc72168 Human NOV  
Acc72167 Human NOV  
Acc72166 Human NOV  
Acc72165 Human NOV  
Acc72164 Human NOV  
Acc72163 Human NOV  
Acc72162 Human NOV  
Acc72161 Human NOV  
Acc72160 Human NOV  
Acc72159 Human NOV  
Acc72158 Human NOV  
Acc72157 Human NOV  
Acc72156 Human NOV  
Acc72155 Human NOV  
Acc72154 Human NOV  
Acc72153 Human NOV  
Acc72152 Human NOV  
Acc72151 Human NOV  
Acc72150 Human NOV  
Acc72149 Human NOV  
Acc72148 Human NOV  
Acc72147 Human NOV  
Acc72146 Human NOV  
Acc72145 Human NOV  
Acc72144 Human NOV  
Acc72143 Human NOV  
Acc72142 Human NOV  
Acc72141 Human NOV  
Acc72140 Human NOV  
Acc72139 Human NOV  
Acc72138 Human NOV  
Acc72137 Human NOV  
Acc72136 Human NOV  
Acc72135 Human NOV  
Acc72134 Human NOV  
Acc72133 Human NOV  
Acc72132 Human NOV  
Acc72131 Human NOV  
Acc72130 Human NOV  
Acc72129 Human NOV  
Acc72128 Human NOV  
Acc72127 Human NOV  
Acc72126 Human NOV  
Acc72125 Human NOV  
Acc72124 Human NOV  
Acc72123 Human NOV  
Acc72122 Human NOV  
Acc72121 Human NOV  
Acc72120 Human NOV  
Acc72119 Human NOV  
Acc72118 Human NOV  
Acc72117 Human NOV  
Acc72116 Human NOV  
Acc72115 Human NOV  
Acc72114 Human NOV  
Acc72113 Human NOV  
Acc72112 Human NOV  
Acc72111 Human NOV  
Acc72110 Human NOV  
Acc72109 Human NOV  
Acc72108 Human NOV  
Acc72107 Human NOV  
Acc72106 Human NOV  
Acc72105 Human NOV  
Acc72104 Human NOV  
Acc72103 Human NOV  
Acc72102 Human NOV  
Acc72101 Human NOV  
Acc72100 Human NOV  
Acc72099 Human NOV  
Acc72098 Human NOV  
Acc72097 Human NOV  
Acc72096 Human NOV  
Acc72095 Human NOV  
Acc72094 Human NOV  
Acc72093 Human NOV  
Acc72092 Human NOV  
Acc72091 Human NOV  
Acc72090 Human NOV  
Acc72089 Human NOV  
Acc72088 Human NOV  
Acc72087 Human NOV  
Acc72086 Human NOV  
Acc72085 Human NOV  
Acc72084 Human NOV  
Acc72083 Human NOV  
Acc72082 Human NOV  
Acc72081 Human NOV  
Acc72080 Human NOV  
Acc72079 Human NOV  
Acc72078 Human NOV  
Acc72077 Human NOV  
Acc72076 Human NOV  
Acc72075 Human NOV  
Acc72074 Human NOV  
Acc72073 Human NOV  
Acc72072 Human NOV  
Acc72071 Human NOV  
Acc72070 Human NOV  
Acc72069 Human NOV  
Acc72068 Human NOV  
Acc72067 Human NOV  
Acc72066 Human NOV  
Acc72065 Human NOV  
Acc72064 Human NOV  
Acc72063 Human NOV  
Acc72062 Human NOV  
Acc72061 Human NOV  
Acc72060 Human NOV  
Acc72059 Human NOV  
Acc72058 Human NOV  
Acc72057 Human NOV  
Acc72056 Human NOV  
Acc72055 Human NOV  
Acc72054 Human NOV  
Acc72053 Human NOV  
Acc72052 Human NOV  
Acc72051 Human NOV  
Acc72050 Human NOV  
Acc72049 Human NOV  
Acc72048 Human NOV  
Acc72047 Human NOV  
Acc72046 Human NOV  
Acc72045 Human NOV  
Acc72044 Human NOV  
Acc72043 Human NOV  
Acc72042 Human NOV  
Acc72041 Human NOV  
Acc72040 Human NOV  
Acc72039 Human NOV  
Acc72038 Human NOV  
Acc72037 Human NOV  
Acc72036 Human NOV  
Acc72035 Human NOV  
Acc72034 Human NOV  
Acc72033 Human NOV  
Acc72032 Human NOV  
Acc72031 Human NOV  
Acc72030 Human NOV  
Acc72029 Human NOV  
Acc72028 Human NOV  
Acc72027 Human NOV  
Acc72026 Human NOV  
Acc72025 Human NOV  
Acc72024 Human NOV  
Acc72023 Human NOV  
Acc72022 Human NOV  
Acc72021 Human NOV  
Acc72020 Human NOV  
Acc72019 Human NOV  
Acc72018 Human NOV  
Acc72017 Human NOV  
Acc72016 Human NOV  
Acc72015 Human NOV  
Acc72014 Human NOV  
Acc72013 Human NOV  
Acc72012 Human NOV  
Acc72011 Human NOV  
Acc72010 Human NOV  
Acc72009 Human NOV  
Acc72008 Human NOV  
Acc72007 Human NOV  
Acc72006 Human NOV  
Acc72005 Human NOV  
Acc72004 Human NOV  
Acc72003 Human NOV  
Acc72002 Human NOV  
Acc72001 Human NOV  
Acc72000 Human NOV  
Acc71999 Human NOV  
Acc71998 Human NOV  
Acc71997 Human NOV  
Acc71996 Human NOV  
Acc71995 Human NOV  
Acc71994 Human NOV  
Acc71993 Human NOV  
Acc71992 Human NOV  
Acc71991 Human NOV  
Acc71990 Human NOV  
Acc71989 Human NOV  
Acc71988 Human NOV  
Acc71987 Human NOV  
Acc71986 Human NOV  
Acc71985 Human NOV  
Acc71984 Human NOV  
Acc71983 Human NOV  
Acc71982 Human NOV  
Acc71981 Human NOV  
Acc71980 Human NOV  
Acc71979 Human NOV  
Acc71978 Human NOV  
Acc71977 Human NOV  
Acc71976 Human NOV  
Acc71975 Human NOV  
Acc71974 Human NOV  
Acc71973 Human NOV  
Acc71972 Human NOV  
Acc71971 Human NOV  
Acc71970 Human NOV  
Acc71969 Human NOV  
Acc71968 Human NOV  
Acc71967 Human NOV  
Acc71966 Human NOV  
Acc71965 Human NOV  
Acc71964 Human NOV  
Acc71963 Human NOV  
Acc71962 Human NOV  
Acc71961 Human NOV  
Acc71960 Human NOV  
Acc71959 Human NOV  
Acc71958 Human NOV  
Acc71957 Human NOV  
Acc71956 Human NOV  
Acc71955 Human NOV  
Acc71954 Human NOV  
Acc71953 Human NOV  
Acc71952 Human NOV  
Acc71951 Human NOV  
Acc71950 Human NOV  
Acc71949 Human NOV  
Acc71948 Human NOV  
Acc71947 Human NOV  
Acc71946 Human NOV  
Acc71945 Human NOV  
Acc71944 Human NOV  
Acc71943 Human NOV  
Acc71942 Human NOV  
Acc71941 Human NOV  
Acc71940 Human NOV  
Acc71939 Human NOV  
Acc71938 Human NOV  
Acc71937 Human NOV  
Acc71936 Human NOV  
Acc71935 Human NOV  
Acc71934 Human NOV  
Acc71933 Human NOV  
Acc71932 Human NOV  
Acc71931 Human NOV  
Acc71930 Human NOV  
Acc71929 Human NOV  
Acc71928 Human NOV  
Acc71927 Human NOV  
Acc71926 Human NOV  
Acc71925 Human NOV  
Acc71924 Human NOV  
Acc71923 Human NOV  
Acc71922 Human NOV  
Acc71921 Human NOV  
Acc71920 Human NOV  
Acc71919 Human NOV  
Acc71918 Human NOV  
Acc71917 Human NOV  
Acc71916 Human NOV  
Acc71915 Human NOV  
Acc71914 Human NOV  
Acc71913 Human NOV  
Acc71912 Human NOV  
Acc71911 Human NOV  
Acc71910 Human NOV  
Acc71909 Human NOV  
Acc71908 Human NOV  
Acc71907 Human NOV  
Acc71906 Human NOV  
Acc71905 Human NOV  
Acc71904 Human NOV  
Acc71903 Human NOV  
Acc71902 Human NOV  
Acc71901 Human NOV  
Acc71900 Human NOV  
Acc71899 Human NOV  
Acc71898 Human NOV  
Acc71897 Human NOV  
Acc71896 Human NOV  
Acc71895 Human NOV  
Acc71894 Human NOV  
Acc71893 Human NOV  
Acc71892 Human NOV  
Acc71891 Human NOV  
Acc71890 Human NOV  
Acc71889 Human NOV  
Acc71888 Human NOV  
Acc71887 Human NOV  
Acc71886 Human NOV  
Acc71885 Human NOV  
Acc71884 Human NOV  
Acc71883 Human NOV  
Acc71882 Human NOV  
Acc71881 Human NOV  
Acc71880 Human NOV  
Acc71879 Human NOV  
Acc71878 Human NOV  
Acc71877 Human NOV  
Acc71876 Human NOV  
Acc71875 Human NOV  
Acc71874 Human NOV  
Acc71873 Human NOV  
Acc71872 Human NOV  
Acc71871 Human NOV  
Acc71870 Human NOV  
Acc71869 Human NOV  
Acc71868 Human NOV  
Acc71867 Human NOV  
Acc71866 Human NOV  
Acc71865 Human NOV  
Acc71864 Human NOV  
Acc71863 Human NOV  
Acc71862 Human NOV  
Acc71861 Human NOV  
Acc71860 Human NOV  
Acc71859 Human NOV  
Acc71858 Human NOV  
Acc71857 Human NOV  
Acc71856 Human NOV  
Acc71855 Human NOV  
Acc71854 Human NOV  
Acc71853 Human NOV  
Acc71852 Human NOV  
Acc71851 Human NOV  
Acc71850 Human NOV  
Acc71849 Human NOV  
Acc71848 Human NOV  
Acc71847 Human NOV  
Acc71846 Human NOV  
Acc71845 Human NOV  
Acc71844 Human NOV  
Acc71843 Human NOV  
Acc71842 Human NOV  
Acc71841 Human NOV  
Acc71840 Human NOV  
Acc71839 Human NOV  
Acc71838 Human NOV  
Acc71837 Human NOV  
Acc71836 Human NOV  
Acc71835 Human NOV  
Acc71834 Human NOV  
Acc71833 Human NOV  
Acc71832 Human NOV  
Acc71831 Human NOV  
Acc71830 Human NOV  
Acc71829 Human NOV  
Acc71828 Human NOV  
Acc71827 Human NOV  
Acc71826 Human NOV  
Acc71825 Human NOV  
Acc71824 Human NOV  
Acc71823 Human NOV  
Acc71822 Human NOV  
Acc71821 Human NOV  
Acc71820 Human NOV  
Acc71819 Human NOV  
Acc71818 Human NOV  
Acc71817 Human NOV  
Acc71816 Human NOV  
Acc71815 Human NOV  
Acc71814 Human NOV  
Acc71813 Human NOV  
Acc71812 Human NOV  
Acc71811 Human NOV  
Acc71810 Human NOV  
Acc71809 Human NOV  
Acc71808 Human NOV  
Acc71807 Human NOV  
Acc71806 Human NOV  
Acc71805 Human NOV  
Acc71804 Human NOV  
Acc71803 Human NOV  
Acc71802 Human NOV  
Acc71801 Human NOV  
Acc71800 Human NOV  
Acc71799 Human NOV  
Acc71798 Human NOV  
Acc71797 Human NOV  
Acc71796 Human NOV  
Acc71795 Human NOV  
Acc71794 Human NOV  
Acc71793 Human NOV  
Acc71792 Human NOV  
Acc71791 Human NOV  
Acc71790 Human NOV  
Acc71789 Human NOV  
Acc71788 Human NOV  
Acc71787 Human NOV  
Acc71786 Human NOV  
Acc71785 Human NOV  
Acc71784 Human NOV  
Acc71783 Human NOV  
Acc71782 Human NOV  
Acc71781 Human NOV  
Acc71780 Human NOV  
Acc71779 Human NOV  
Acc71778 Human NOV  
Acc71777 Human NOV  
Acc71776 Human NOV  
Acc71775 Human NOV  
Acc71774 Human NOV  
Acc71773 Human NOV  
Acc71772 Human NOV  
Acc71771 Human NOV  
Acc71770 Human NOV  
Acc71769 Human NOV  
Acc71768 Human NOV  
Acc71767 Human NOV  
Acc71766 Human NOV  
Acc71765 Human NOV  
Acc71764 Human NOV  
Acc71763 Human NOV  
Acc71762 Human NOV  
Acc71761 Human NOV  
Acc71760 Human NOV  
Acc71759 Human NOV  
Acc71758 Human NOV  
Acc71757 Human NOV  
Acc71756 Human NOV  
Acc71755 Human NOV  
Acc71754 Human NOV  
Acc71753 Human NOV  
Acc71752 Human NOV  
Acc71751 Human NOV  
Acc71750 Human NOV  
Acc71749 Human NOV  
Acc71748 Human NOV  
Acc71747 Human NOV  
Acc71746 Human NOV  
Acc71745 Human NOV  
Acc71744 Human NOV  
Acc71743 Human NOV  
Acc71742 Human NOV  
Acc71741 Human NOV  
Acc71740 Human NOV  
Acc71739 Human NOV  
Acc71738 Human NOV  
Acc71737 Human NOV  
Acc71736 Human NOV  
Acc71735 Human NOV  
Acc71734 Human NOV  
Acc71733 Human NOV  
Acc71732 Human NOV  
Acc71731 Human NOV  
Acc71730 Human NOV  
Acc71729 Human NOV  
Acc71728 Human NOV  
Acc71727 Human NOV  
Acc71726 Human NOV  
Acc71725 Human NOV  
Acc71724 Human NOV  
Acc71723 Human NOV  
Acc71722 Human NOV  
Acc71721 Human NOV  
Acc71720 Human NOV  
Acc71719 Human NOV  
Acc71718 Human NOV  
Acc71717 Human NOV  
Acc71716 Human NOV  
Acc71715 Human NOV  
Acc71714 Human NOV  
Acc71713 Human NOV  
Acc71712 Human NOV  
Acc71711 Human NOV  
Acc71710 Human NOV  
Acc71709 Human NOV  
Acc71708 Human NOV  
Acc71707 Human NOV  
Acc71706 Human NOV  
Acc71705 Human NOV  
Acc71704 Human NOV  
Acc71703 Human NOV  
Acc71702 Human NOV  
Acc71701 Human NOV  
Acc71700 Human NOV  
Acc71699 Human NOV  
Acc71698 Human NOV  
Acc71697 Human NOV  
Acc71696 Human NOV  
Acc71695 Human NOV  
Acc71694 Human NOV  
Acc71693 Human NOV  
Acc71692 Human NOV  
Acc71691 Human NOV  
Acc71690 Human NOV  
Acc71689 Human NOV  
Acc71688 Human NOV  
Acc71687 Human NOV  
Acc71686 Human NOV  
Acc71685 Human NOV  
Acc71684 Human NOV  
Acc71683 Human NOV  
Acc71682 Human NOV  
Acc71681 Human NOV  
Acc71680 Human NOV  
Acc71679 Human NOV  
Acc71678 Human NOV  
Acc71677 Human NOV  
Acc71676 Human NOV  
Acc71675 Human NOV  
Acc71674 Human NOV  
Acc71673 Human NOV  
Acc71672 Human NOV  
Acc71671 Human NOV  
Acc71670 Human NOV  
Acc71669 Human NOV  
Acc71668 Human NOV  
Acc71667 Human NOV  
Acc71666 Human NOV  
Acc71665 Human NOV  
Acc71664 Human NOV  
Acc71663 Human NOV  
Acc71662 Human NOV  
Acc71661 Human NOV  
Acc71660 Human NOV  
Acc71659 Human NOV  
Acc71658 Human NOV  
Acc71657 Human NOV  
Acc71656 Human NOV  
Acc71655 Human NOV  
Acc71654 Human NOV  
Acc71653 Human NOV  
Acc71652 Human NOV  
Acc71651 Human NOV  
Acc71650 Human NOV  
Acc71649 Human NOV  
Acc71648 Human NOV  
Acc71647 Human NOV  
Acc71646 Human NOV  
Acc71645 Human NOV  
Acc71644 Human NOV  
Acc71643 Human NOV  
Acc71642 Human NOV  
Acc71641 Human NOV  
Acc71640 Human NOV  
Acc71639 Human NOV  
Acc71638 Human NOV  
Acc71637 Human NOV  
Acc71636 Human NOV  
Acc71635 Human NOV  
Acc71634 Human NOV  
Acc71633 Human NOV  
Acc71632 Human NOV  
Acc71631 Human NOV  
Acc71630 Human NOV  
Acc71629 Human NOV  
Acc71628 Human NOV  
Acc71627 Human NOV  
Acc71626 Human NOV  
Acc71625 Human NOV  
Acc71624 Human NOV  
Acc71623 Human NOV  
Acc71622 Human NOV  
Acc71621 Human NOV  
Acc71620 Human NOV  
Acc71619 Human NOV  
Acc71618 Human NOV  
Acc71617 Human NOV  
Acc71616 Human NOV  
Acc71615 Human NOV  
Acc71614 Human NOV  
Acc71613 Human NOV  
Acc71612 Human NOV  
Acc71611 Human NOV  
Acc71610 Human NOV  
Acc71609 Human NOV  
Acc71608 Human NOV  
Acc71607 Human NOV  
Acc71606 Human NOV  
Acc71605 Human NOV  
Acc71604 Human NOV  
Acc71603 Human NOV  
Acc71602 Human NOV  
Acc71601 Human NOV  
Acc71600 Human NOV  
Acc71599 Human NOV  
Acc71598 Human NOV  
Acc71597 Human NOV  
Acc71596 Human NOV  
Acc71595 Human NOV  
Acc71594 Human NOV  
Acc71593 Human NOV  
Acc71592 Human NOV  
Acc71591 Human NOV  
Acc71590 Human NOV  
Acc71589 Human NOV  
Acc71588 Human NOV  
Acc71587 Human NOV  
Acc71586 Human NOV  
Acc71585 Human NOV  
Acc71584 Human NOV  
Acc71583 Human NOV  
Acc71582 Human NOV  
Acc71581 Human NOV  
Acc71580 Human NOV  
Acc71579 Human NOV  
Acc71578 Human NOV  
Acc71577 Human NOV  
Acc71576 Human NOV  
Acc71575 Human NOV  
Acc71574 Human NOV  
Acc71573 Human NOV  
Acc71572 Human NOV  
Acc71571 Human NOV  
Acc71570 Human NOV  
Acc71569 Human NOV  
Acc71568 Human NOV  
Acc71567 Human NOV  
Acc71566 Human NOV  
Acc71565 Human NOV  
Acc71564 Human NOV  
Acc71563 Human NOV  
Acc71562 Human NOV  
Acc71561 Human NOV  
Acc71560 Human NOV  
Acc71559 Human NOV  
Acc71558 Human NOV  
Acc71557 Human NOV  
Acc71556 Human NOV  
Acc71555 Human NOV  
Acc71554 Human NOV  
Acc71553 Human NOV  
Acc71552 Human NOV  
Acc71551 Human NOV  
Acc71550 Human NOV  
Acc71549 Human NOV  
Acc71548 Human NOV  
Acc71547 Human NOV  
Acc71546 Human NOV  
Acc71545 Human NOV  
Acc71544 Human NOV  
Acc71543 Human NOV  
Acc71542 Human NOV  
Acc71541 Human NOV  
Acc71540 Human NOV  
Acc71539 Human NOV  
Acc71538 Human NOV  
Acc71537 Human NOV  
Acc71536 Human NOV  
Acc71535 Human NOV  
Acc71534 Human NOV  
Acc71533 Human NOV  
Acc71532 Human NOV  
Acc71531 Human NOV  
Acc71530 Human NOV  
Acc71529 Human NOV  
Acc71528 Human NOV  
Acc71527 Human NOV  
Acc71526 Human NOV  
Acc71525 Human NOV  
Acc71524 Human NOV  
Acc71523 Human NOV  
Acc71522 Human NOV  
Acc71521 Human NOV  
Acc71520 Human NOV  
Acc71519 Human NOV  
Acc71518 Human NOV  
Acc71517 Human NOV  
Acc71516 Human NOV  
Acc71515 Human NOV  
Acc71514 Human NOV  
Acc71513 Human NOV  
Acc71512 Human NOV  
Acc71511 Human NOV  
Acc71510 Human NOV  
Acc71509 Human NOV  
Acc71508 Human NOV  
Acc71507 Human NOV  
Acc71506 Human NOV  
Acc71505 Human NOV  
Acc71504 Human NOV  
Acc71503 Human NOV  
Acc71502 Human NOV  
Acc71501 Human NOV  
Acc71500 Human NOV  
Acc71499 Human NOV  
Acc71498 Human NOV  
Acc71497 Human NOV  
Acc71496 Human NOV  
Acc71495 Human NOV  
Acc71494 Human NOV  
Acc71493 Human NOV  
Acc71492 Human NOV  
Acc71491 Human NOV  
Acc71490 Human NOV  
Acc71489 Human NOV  
Acc71488 Human NOV  
Acc71487 Human NOV  
Acc71486 Human NOV  
Acc71485 Human NOV  
Acc71484 Human NOV  
Acc71483 Human NOV  
Acc71482 Human NOV  
Acc71481 Human NOV  
Acc71480 Human NOV  
Acc71479 Human NOV  
Acc71478 Human NOV  
Acc71477 Human NOV  
Acc71476 Human NOV  
Acc71475 Human NOV  
Acc71474 Human NOV  
Acc71473 Human NOV  
Acc71472 Human NOV  
Acc71471 Human NOV  
Acc71470 Human NOV  
Acc71469 Human NOV  
Acc71468 Human NOV  
Acc71467 Human NOV  
Acc71466 Human NOV  
Acc71465 Human NOV  
Acc71464 Human NOV  
Acc71463 Human NOV  
Acc71462 Human NOV  
Acc71461 Human NOV  
Acc71460 Human NOV  
Acc71459 Human NOV  
Acc71458 Human NOV  
Acc71457 Human NOV  
Acc71456 Human NOV  
Acc71455 Human NOV  
Acc71454 Human NOV  
Acc71453 Human NOV  
Acc71452 Human NOV  
Acc71451 Human NOV  
Acc71450 Human NOV  
Acc71449 Human NOV  
Acc71448 Human NOV  
Acc71447 Human NOV  
Acc71446 Human NOV  
Acc71445 Human NOV  
Acc71444 Human NOV  
Acc71443 Human NOV  
Acc71442 Human NOV  
Acc71441 Human NOV  
Acc71440 Human NOV  
Acc71439 Human NOV  
Acc71438 Human NOV  
Acc71437 Human NOV  
Acc71436 Human NOV  
Acc71435 Human NOV  
Acc71434 Human NOV  
Acc71433 Human NOV  
Acc71432 Human NOV  
Acc71431 Human NOV  
Acc71430 Human NOV  
Acc71429 Human NOV  
Acc71428 Human NOV  
Acc71427 Human NOV  
Acc71426 Human NOV  
Acc71425 Human NOV  
Acc71424 Human NOV  
Acc71423 Human NOV  
Acc71422 Human NOV  
Acc71421 Human NOV  
Acc71420 Human NOV  
Acc71419 Human NOV  
Acc71418 Human NOV  
Acc71417 Human NOV  
Acc71416 Human NOV  
Acc71415 Human



CC proteins. The DNA and protein sequences of the invention are useful as:  
CC markers for tissues in which the corresponding protein is preferentially  
CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present DNA sequence represents a gene of the  
CC invention.

XX	SQ	Sequence	5847 BP; 1607 A; 1282 C; 1453 G; 1505 T; 0 U; 0 Other;
		Query Match	99.6%; Score 4380.4; DB 9; Length 5847;
		Best Local Similarity	99.8%; Pred. No. 0;
		Matches 4397; Conservative	0; Mismatches 1; Indels 6; Gaps 1;
QY		1	ATGGTGCCTCGATTCAGAACTGAAACCGAAGCTGTGGTTGCAAAATGAAATGCGAACAC 60
DB		322	ATGGTGCCTCGATTCAGAACTGAAACCGAAGCTGTGGTTGCAAAATGAAATGCGAACAC 381
QY		61	TTTGGCCCTCAGCCTCCACAGAACAGGCTTGCCTCATTCCTTCCCGGGATTGTGTA 120
DB		382	TTTGGCCCTCAGCCTCCACAGAACAGGCTTGCCTCATTCCTTCCCGGGATTGTGTA 441
QY		121	GTATCTGAGTTCTTACCATGTCCAACTGTAGCAAGGAGATGTGGAAAGAAATTCAGCAT 180
DB		442	GTATCTGAGTTCTTACCATGTCCAACTGTAGCAAGGAGATGTGGAAAGAAATTCAGCAT 501
QY		181	AGAACTCCGCGGTGATAGCTCCCGCTCTCTTGGTGGTTGCAATCTCCAAATCTGACT 240
DB		502	AGAACTCCGCGGTGATAGCTCCCGCTCTCTTGGTGGTTGCAATCTCCAAATCTGACT 561
QY		241	GAGTCAAGAGCCTGTGATGCTCCAAATTCCTGTCTCTTGGGGAAGAGGAATATACATTT 300
DB		562	GAGTCAAGAGCCTGTGATGCTCCAAATTCCTGTCTCTTGGGGAAGAGGAATATACATTT 621
QY		301	AGCCTTAGGTTGGACCATGAGTAATGCACTGCTCATCTTAAGAAATTAATCCA 360
DB		622	AGCCTTAGGTTGGACCATGAGTAATGCACTGCTCATCTTAAGAAATTAATCCA 681
QY		361	ACCGGAAGAACTGTTCTGGATTTTAACTCTGATTCAAATGAGCGAGTCACCTTTAAACAT 420
DB		682	ACCGGAAGAACTGTTCTGGATTTTAACTCTGATTCAAATGAGCGAGTCACCTTTAAACAT 741
QY		421	CAAAAGTTACAAAGACATCATCATTCGAAAGTCTTGGGCAATAGAGATAGGTATCAAAAC 480
DB		742	CAAAAGTTACAAAGACATCATCATTCGAAAGTCTTGGGCAATAGAGATAGGTATCAAAAC 801
QY		481	CGGAGAGTTTCGTGTACAAAGATGATGACAAATGCTATGTTAAGCCTTTGCTTCAA 540
DB		802	CGGAGAGTTTCGTGTACAAAGATGATGACAAATGCTATGTTAAGCCTTTGCTTCAA 861
QY		541	GATTCTTCCCATGACATGTTTCAGTCTCTGATCATGCCCAAGACTGTGAAACCTCCAG 600
DB		862	GATTCTTCCCATGACATGTTTCAGTCTCTGATCATGCCCAAGACTGTGAAACCTCCAG 921
QY		601	TGGTCTCTCGAGCCCTGCTCCAAAGACATGCCGTTTCAGGAGTCTTTCGCAAGATTT 660
DB		922	TGGTCTCTCGAGCCCTGCTCCAAAGACATGCCGTTTCAGGAGTCTTTCGCAAGATTT 981
QY		661	AGGAGCAGAGCCGGAACGTGAAACATATGCTATTCGAGGTGAAAGAGTGTCTCTGAA 720
DB		982	AGGAGCAGAGCCGGAACGTGAAACATATGCTATTCGAGGTGAAAGAGTGTCTCTGAA 1041
QY		721	CTTCTTGGAAGAGGCTGCTGATTTGTTGAGGAGAACTTTCGAGCAATCTCCAGGTAT 780
DB		1042	CTTCTTGGAAGAGGCTGCTGATTTGTTGAGGAGAACTTTCGAGCAATCTCCAGGTAT 1101
QY		781	TCCTGGAGAACTTCTGAAATGAAAGAAATGCCAAGTCTCTCTCTCTCGAGCAGCAGGAT 840
DB		1102	TCCTGGAGAACTTCTGAAATGAAAGAAATGCCAAGTCTCTCTCTCTCGAGCAGCAGGAT 1161
QY		841	CCCCACTGGCATGTGACGGGACCGGTGTGGTGGGATCCAGACCCCGGGAGGTGTAC 900
DB		1162	CCCCACTGGCATGTGACGGGACCGGTGTGGTGGGATCCAGACCCCGGGAGGTGTAC 1221

QY	901	TGTGCCAGAGGCTTACCAGAGCTGCCGCACTGAGGCGCAAGGAAGTCTCTAGACCTGTG 960
DB	1222	TGTGCCAGAGGCTTACCAGAGCTGCCGCACTGAGGCGCAAGGAAGTCTCTAGACCTGTG 1281
QY	961	GAAAAGGCATTAATGTGTGGGACCCCGCGTGGTCCCTCTCAGCTCTGCAATATCCCTTGC 1020
DB	1282	GAAAAGGCATTAATGTGTGGGACCCCGCGTGGTCCCTCTCAGCTCTGCAATATCCCTTGC 1341
QY	1021	TCTACGAGCTGCATAGTATCTTCCCTGGTTCAGCTGGGCTGTGTCATCCATGAAAATGT 1080
DB	1342	TCTACGAGCTGCATAGTATCTTCCCTGGTTCAGCTGGGCTGTGTCATCCATGAAAATGT 1401
QY	1081	CATGAACCTCAGGGGAAATAAGGATTTAGACAGGAGCGCCATGTCTCTCATGAAATCT 1140
DB	1402	CATGAACCTCAGGGGAAATAAGGATTTAGACAGGAGCGCCATGTCTCTCATGAAATCT 1461
QY	1141	ACAGGCGCTGCAAGGCATTTGCCCTCATTTGGTGGAGTCTGTTCCTTGTGAGATCCAATG 1200
DB	1462	ACAGGCGCTGCAAGGCATTTGCCCTCATTTGGTGGAGTCTGTTCCTTGTGAGATCCAATG 1521
QY	1201	TGCTACCGATGCTGGCATCAGAGGAGTCTGTTCCTGTATCATGGAATGTGGCCTG 1260
DB	1522	TGCTACCGATGCTGGCATCAGAGGAGTCTGTTCCTGTATCATGGAATGTGGCCTG 1581
QY	1261	GGACATCGTATTTCAAGCGCTCTGCCAGATGACCGCGAGAGAGTATCAGGAGT 1320
DB	1582	GGACATCGTATTTCAAGCGCTCTGCCAGATGACCGCGAGAGAGTATCAGGAGT 1641
QY	1321	CTTTGCCAGTTCCCTCTCTCAGAGGAGTCTTGTGAAATTTCCCTGCCGAATGAC 1380
DB	1642	CTTTGCCAGTTCCCTCTCTCAGAGGAGTCTTGTGAAATTTCCCTGCCGAATGAC 1701
QY	1381	TGCTGTCTGAGCGAGTGGACGAGTGTCTCTCTGTTCCAGTCTCTTCAATAAATAAC 1440
DB	1702	TGCTGTCTGAGCGAGTGGACGAGTGTCTCTCTGTTCCAGTCTCTTCAATAAATAAC 1761
QY	1441	TCAGATGGGAAACAGACACAGCTCAAGAACTATCTCTGGCACTGGCTGGGGAAGTGGAAAG 1500
DB	1762	TCAGATGGGAAACAGACACAGCTCAAGAACTATCTCTGGCACTGGCTGGGGAAGTGGAAAG 1821
QY	1501	CCATGTCCCTTAGTCAGGCTCTCAAGAGCATCTGTTGTGTATGACCATCTCTGTATG 1560
DB	1822	CCATGTCCCTTAGTCAGGCTCTCAAGAGCATCTGTTGTGTATGACCATCTCTGTATG 1881
QY	1561	CAGCTTCACTGGGAGACATCGCTTGGGCGCTTCTCTGAGGACACATTTGGTAATGCC 1620
DB	1882	CAGCTTCACTGGGAGACATCGCTTGGGCGCTTCTCTGAGGACACATTTGGTAATGCC 1941
QY	1621	CTTAATGCAACCATTTGGCTGGAATGGAGAGCCAGTGTGTGTAGGCATTCAGACTGG 1680
DB	1942	CTTAATGCAACCATTTGGCTGGAATGGAGAGCCAGTGTGTGTAGGCATTCAGACTGG 2001
QY	1681	AGAGTCTTCTGTCTCAAGAGTCACTGGGGAACAAGTAATGACCAAAAGATGTCCAGATTCT 1740
DB	2002	AGAGTCTTCTGTCTCAAGAGTCACTGGGGAACAAGTAATGACCAAAAGATGTCCAGATTCT 2061
QY	1741	ACTGCACTGAAACTGTGCGCCCTCTGTTTCTCCATGCAAAAAGACTGTATTGTGACT 1800
DB	2062	ACTGCACTGAAACTGTGCGCCCTCTGTTTCTCCATGCAAAAAGACTGTATTGTGACT 2121
QY	1801	GCTTTTCAGTGTGACACCTCTGCCAAGGATGTGCCAAGCAGGAAATGCCACAGTAAA 1860
DB	2122	GCTTTTCAGTGTGACACCTCTGCCAAGGATGTGCCAAGCAGGAAATGCCACAGTAAA 2181
QY	1861	CAGTCTCGATACAGAAATCATCATCAAGAGCAGCAATGGAGGCGCAGGAATGCCAGAT 1920
DB	2182	CAGTCTCGATACAGAAATCATCATCAAGAGCAGCAATGGAGGCGCAGGAATGCCAGAT 2241
QY	1921	ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTGTGTCTGTATATCGGTGGAG 1980
DB	2242	ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTGTGTCTGTATATCGGTGGAG 2301



Db 4462 GCTGATGTGAAAAACCTTTCTGTGGAAAAACAGACCTGTGTAATTCAAAAATACATGATATT 4521  
QY 4195 TTTAAAGGATGGTCTCTTCAACCACTTGATCCAGATGGCGAGTAAATAATTTGGGTTTAT 4254  
Db 4522 TTTAAAGGATGGTCTCTTCAACCACTTGATCCAGATGGCGAGTAAATAATTTGGGTTTAT 4581  
QY 4255 GGCCTTTGAGTGGCGCTTTCTCATCATGATTTTCTTAATATTACTTCTACCTTGTT 4314  
Db 4582 GGCCTTTGAGTGGCGCTTTCTCATCATGATTTTCTTAATATTACTTCTACCTTGTT 4641  
QY 4315 TGCAGAGCCAAACACATCAAGACACACCTCCCAACAGAGCCTCTGACCTTAGCC 4374  
Db 4642 TGCAGAGCCAAACACATCAAGACACACCTCCCAACAGAGCCTCTGACCTTAGCC 4701  
QY 4375 TACGATGGAGACTTAGACATGTAA 4398  
Db 4702 TACGATGGAGACTTAGACATGTAA 4725

## RESULT 2

AAL43703

ID AAL43703 standard; cDNA; 3382 BP.

XX

AC

XX

AC

XX

DT

XX

XX

DE

XX

XX

KW

XX

KW

XX

OS

XX

XX

FH

XX

FT

XX

FT

XX

FT

XX

XX

PN

XX

XX

PD

XX

XX

PF

XX

XX

PR

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

16-SEP-2002 (first entry)

Human large protein 53-24 cDNA sequence.

Human; gene; ss; large protein 53.24; embryonic development deformity;  
protein metabolism disturbance; tumour; immunologic system disturbance.

Homo sapiens.

Key Location/Qualifiers

CDS 819..2273

/\*tag= a

/product= "Human large protein 53-24"

CNI133246-A.

30-JAN-2002.

07-JUL-2000; 2000CN-00117029.

07-JUL-2000; 2000CN-00117029.

(SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.

Mao Y, Xie Y;

WPI; 2002-305572/35.

P-PSDB; AAC15279.

New human large protein 53.24 and encoding polynucleotide, useful for  
treating embryonic development deformity, protein metabolism disturbance,  
tumor and immunologic system disturbance disease.

Claim 6; Page 27-29 (Disclosure); 36pp; Chinese.

The invention comprises the amino acid and coding sequence of the human  
large protein 53.24. The 53.24 DNA and protein sequences of the invention  
are useful for treating embryonic development deformity, protein  
metabolism disturbance, tumour and immunologic system disturbance. The  
present cDNA sequence encodes the human large protein 53.24

Sequence 3382 BP; 983 A; 705 C; 779 G; 915 T; 0 U; 0 Other;

Query Match

Best Local Similarity 50.8%; Score 2235.8; DB 6; Length 3382;

Matches 2264; Conservative 0; Mismatches 2; Indels 7; Gaps 2;

QY 2133 GCAGACAAACGGCAGCTCTCTCTGTGCAAGAAATGCACAGTCCCATGTGAGAGAACTG 2192

Db 1 GCAGACAAACGGCAGCTCTCTCTGTGCAAGAAATGCACAGTCCCATGTGAGAGAACTG 60  
QY 2193 CACCTTCACTGCTGGTCCAAAGTTTACGCCCTGCTCCACGAACTGTGAGACCAAAAG 2252  
Db 61 CACCTTCACTGCTGGTCCAAAGTTTACGCCCTGCTCCACGAACTGTGAGACCAAAAG 120  
QY 2253 TAGGCGGCGACAGCTCACAGGAAAAAGCAGAAAAAGAGAGAAATGCCAGGATTTCTGACCT 2312  
Db 121 TAGGCGGCGACAGCTCACAGGAAAAAGCAGAAAAAGAGAGAAATGCCAGGATTTCTGACCT 180  
QY 2313 TTACCTCTAGTGGAGACAGAACTATGCTTGTGATGAAATTTATATCCCAACCTTATGG 2372  
Db 181 TTACCTCTAGTGGAGACAGAACTATGCTTGTGATGAAATTTATATCCCAACCTTATGG 240  
QY 2373 AAACTGGTCAAGTTGCAATTTCTTCAGAAAGGAGAGGAGCTCACCGAGACTGCGGGT 2432  
Db 241 AAACTGGTCAAGTTGCAATTTCTTCAGAAAGGAGAGGAGGAGCTCACCGAGACTGCGGGT 300  
QY 2433 ACAAGCAGACAGCAAGAAATGTGGAGAAAGGCTGGCTTTTCGAGAGTAGCTGTTCTGA 2492  
Db 301 ACAAGCAGACAGCAAGAAATGTGGAGAAAGGCTGGCTTTTCGAGAGTAGCTGTTCTGA 360  
QY 2493 TAAAAATGCAAGACCTGTTGACCCCTCTCTGAGAGAGCTCTGTTACATTCAGAAAA 2552  
Db 361 TAAAAATGCAAGACCTGTTGACCCCTCTCTGAGAGAGCTCTGTTACATTCAGAAAA 420  
QY 2553 ATGTCTCATTCCTGCCCATTTGATTTGCAAGTTAAGCAATGCTGTTAGTTGGGGTCTTG 2612  
Db 421 ATGTCTCATTCCTGCCCATTTGATTTGCAAGTTAAGCAATGCTGTTAGTTGGGGTCTTG 480  
QY 2613 CAGTTTCATCTTGTGAAATGGAGTGAGAAATTCGATCCAAATGGCTT-AAAAAGAAAAACCTT 2671  
Db 481 CAGTTTCATCTTGTGAAATGGAGTGAGAAATTCGATCCAAATGGCTT-AAAAAGAAAAACCTT 540  
QY 2672 ACAATGGAGGACGACCATGTCCAAACTGGATCTCAAGAA-----TCAGGTACATGAGG 2725  
Db 541 ACAATGGAGGACGACCATGTCCAAACTGGATCTCAAGAACTCAGGCTCAGGTACATGAGG 600  
QY 2726 CAGTCCCATGTTACAGTGAGTGCAATCAGTATCTCGGTTGTAGAACACTGCTTCTCAT 2785  
Db 601 CAGTCCCATGTTACAGTGAGTGCAATCAGTATCTCGGTTGTAGAACACTGCTTCTCAT 660  
QY 2786 GCAAAATCAACAATGAGCTGAGGTCCTCGCTGCTGGAGGAGGAAACAACAATCTAGAAAA 2845  
Db 661 GCAAAATCAACAATGAGCTGAGGTCCTCGCTGCTGGAGGAGGAGGAAACAACAATCTAGAAAA 720  
QY 2846 TCAGATGTGTAATGAGTCTGGGATGTAAGGTGGAGCAGTGATAGCAACCTGTGCAACC 2905  
Db 721 TCAGATGTGTAATGAGTCTGGGATGTAAGGTGGAGCAGTGATAGCAACCTGTGCAACC 780  
QY 2906 AGGATGAAATTCCTCCAGAAACCCAGTCTGTTCTCTTATGTTGTCCTCCATGAGTGTGCA 2965  
Db 781 AGGATGAAATTCCTCCAGAAACCCAGTCTGTTCTCTTATGTTGTCCTCCATGAGTGTGCA 840  
QY 2966 TGTCTGAGTGGGAACTTTGGAGCAAAATGCCACAGTCAATGGAATCCCAACAATGCGAGA 3025  
Db 841 TGTCTGAGTGGGAACTTTGGAGCAAAATGCCACAGTCAATGGAATCCCAACAATGCGAGA 900  
QY 3026 GAAGAACTCGCCACCTGCTTAAGACCATCACTGAATCAAGGACTTGTCTGAAGACTCAC 3085  
Db 901 GAAGAACTCGCCACCTGCTTAAGACCATCACTGAATCAAGGACTTGTCTGAAGACTCAC 960  
QY 3086 AGGTGAGCCTTGCCTCTCTGAAATGAAATTTGCTCCAGTTCAGTACCAATCTAACAGAGT 3145  
Db 961 AGGTGAGCCTTGCCTCTCTGAAATGAAATTTGCTCCAGTTCAGTACCAATCTAACAGAGT 1020  
QY 3146 GCAGACATGCCAGCTGAGTGAACGACACCTGTGGTCAAGGCGTCAGGACCCGCTGC 3205  
Db 1021 GCAGACATGCCAGCTGAGTGAACGACACCTGTGGTCAAGGCGTCAGGACCCGCTGC 1080  
QY 3206 TAAGCTGTGTGTCAGTGTGAGTGAACGACCATGTCAGTGCACCAATGTGAGCAGCATATT 3265



CC The present invention relates to novel human NOV proteins and their  
CC coding sequences (ACC72075-ACC72181 and AB858469). The NOV  
CC proteins are useful in manufacturing a medicament for treating a syndrome  
CC associated with a human disease. The NOV proteins and coding sequences  
CC may be used to diagnose, treat or prevent metabolic disorders such as  
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
CC disorders such as Alzheimer's disease or Parkinson's disease, immune  
CC disorders, haematopoietic disorders and various dyslipidaemias  
XX  
SQ Sequence 4810 BP; 1351 A; 1104 C; 1317 G; 1038 T; 0 U; 0 Other;

Query Match 28.0%; Score 1232.8; DB 7; Length 4810;  
Best Local Similarity 57.7%; Pred. No. 0;  
Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;  
QY 10 TGCATTGAGAGCTGAACCGAATCTGTTGGCAATGAAATATGCAACTTTGGCCCTT 69  
DB 288 TGCATCAGAAAGACAAGACATCTCTGGAGGATATCATCTGTGAGTACTTTGAGCCC 347  
QY 70 CAGCTCTCTACAGAACAGGTGCTCATCTCTGTCCTCCCGGATTTGTGATCTGAG 129  
DB 348 AAGCCTCTCTGAGACAGGTGCTCATCTCTGTCCTCCAGCAAGATTGCATCGTGTCTGAA 407  
QY 130 TTCTTACATGCTCCCACTGTAGCAAGGATGCGAAGAAATGACGATACAACTCGC 189  
DB 408 TTTTCTGCTGCTCCGAATGCTCAAGACCTGCGGAGCGGCTCCAGCACCGAGCGGT 467  
QY 190 GCGGTCAATGCTCCCTCTCTTTTGTGTTTGCATGTCCTCAATGTCAGTGTCAAGA 249  
DB 468 CATGTGTTGGCGCCCGCCAGTTCGGAGGCTCTGCTGTCCAAACCTGACGGAGTTCAG 527  
QY 250 GCCTGTGATGCTCCCATTTCTCTGCTCTTGGGAGAGGAATATACATTTAGCCTTAAG 309  
DB 528 GTGTG-----CCAACTCAGTCCATGCGAGCGGAGGTTCAGGTACAGCCTTGCAT 578  
QY 310 GTTGGACATGAGTAATGAGACTGCTCCTCATCTTAAAGAAATTAATCCAGCGGAAGA 369  
DB 579 GTGGGCGCTGGACACCTGCTCAATGSCCCACTCCCGACAAGTAAGACAAGCAAGAGA 638  
QY 370 ACTGTTCTGANTTTAA-----CTCTGATTCAAATGAG 402  
DB 639 CGCGGGAAGAAATAAGAACCGGAAAGGACCGCAGCAAGAGGTAAGGATCCAGAGCC 698  
QY 403 CGAGTCACTTTAAACATCAAAGTTACA-----AAGCACATCATCATTCGAAGTCT 453  
DB 699 CGCAGCTTATTAAGAAAGAGAAACAGAAACAGGACAGACAGACAAGACAACATAT 758  
QY 454 TGGCAATAGAGATAGTATTCAACCCCGAGGTTTCGTGTACAGAAAGTGTAGACAA 513  
DB 759 TGGACATCCAGATTGGATATCAGACCAGAGAGGTTATGTGCATTAAACAAGCGGGAAA 818  
QY 514 AATGCTATGTTAGCCCTTGCCTTCAAGATTCCTCCCATTTGACTGTTCAGTCCCTGATC 573  
DB 819 GCTGCTGATTTAAGCTTTTGGCAGCAGAGAGGCTTCCATGACTTCCAGTCTCTGTG 878  
QY 574 ATGCCCAAAGACTGTGAACCTCCAGTGTCTCTGGAGCCCTGCTCCAAAGACATGC 633  
DB 879 ATCAACAAAGAGTCCAGGTTTCGAGTGTGTCAGAGTGGAGCCCTGCTCAAAAACATGC 938  
QY 634 CGTTACGGAGTCTCTTGGCAGGATTTAGAGCAGGAGCCGGAACGTGAAGCATGCGCT 593  
DB 939 CATGACATGTTGCTCCCTGTCAGGACCTCGTGTGAAGGACAGAACCATCAGGAGTTCC 998  
QY 694 ATTGGAGGTGGAAGAGTGTCTCTGAACTTCTTGAAGAGAGGCGCTGCATTTGTGAAGA 753  
DB 999 ATTGGCAGTGAAGAGAGTGTCCAGATTTTGAAGAAAGAAACCTGTTGTCTCAAGGA 1058  
QY 754 GAATCTTCGAGCAATGTCAGGATTTCTGAGAACTTCTGAAATGGAAGAAATGCCAA 813  
DB 1059 GATGGAGTTGTCCCTGTGCAAGTATGAGTGTGAGAACTACAGAGTGGAGTGAAGTGCCT 1118  
QY 814 GTCTCTCTCTCTCCGAGCAGAGGATCCCCACTGGCATGTGACGGGACCGGTGTGTGGC 873

DB 1119 GTGGACCCCTTTGCTCAGTCAGCAGACAGAGGCGCGCAACACAGACGCGCCTCTGTGGA 1178  
QY 874 GGTGGATCCAGACCCGGAGGTGTACTGTGCCAG-----AGGTACACGCA 921  
DB 1179 GGGGCAATCCAGACCCGAGAGGTGTACTGTGTCGAGGCCAACGAAAAACCTCTCTCAAA 1238  
QY 922 GCTCGGCACACTGAGGGCCCAAGAAAGTCTCTAGACCTGTGGAAGGCAATTTATGTGTGGGA 981  
DB 1239 TTAAGTACCCACAAGAACAAAGAGCCTCAAGCAATGGACTTTAAAAATTATGCACTGGA 1298  
QY 982 CCGGCCCGTTGGCCTCTCAGCTCTGCAATATCCCTTGCTCTACGGAATGATATCT 1041  
DB 1299 CCTATCCCTAATATACACAGCTGTGCCAATCTCTTGTCCAACTGAATGTGAAGTTTCA 1358  
QY 1042 TCCTGCTCAGCCTGGGCGCTGTGCATCATGAAACCTGTATGAACCTCAGGGGAAAAAAA 1101  
DB 1359 CTTTGGTCAGCTTGGGACCTTGTATGAAACCTGTATGATCAGCAAGGAAAAAAA 1418  
QY 1102 GGATTTAGAACGAGCAGCG-----CATGTCTCATGGAATCTACAGGSCCTGCA 1152  
DB 1419 GGTTCAAACTGAGAAAGCGCGCAATTACCAATGAGCCCACTGGAGCTCTGGGGTAACC 1478  
QY 1153 GGGCAATTCCTCATTTGCTGGAGTCTGCTTCTGTGAGGATCCAAATGTCTACCCGATGG 1212  
DB 1479 GGAACCTGCTCTACTTCTGGAAGCCATTCCTGTGAGAGCTGCTGTATGACTGG 1538  
QY 1213 ---CTGGCATCAGAAAGGATCTTTCCCTGATCATGAA---AATGTGGCCTGGGACAT 1266  
DB 1539 AAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAGAGGTGTGTCTCCAGGACG 1598  
QY 1267 CGTATTCGAGGCGCTGTCAGCAATGACCGGAGAAAGTATCAGGGAGTCTTTGC 1356  
DB 1599 CAAGTTCAAGAGGTTGTGTGATCAACAGTATGGAAGAAAGTTGACACACACTGTGC 1658  
QY 1327 CCAGTTCCTCTCTCTGAGAGGAGTCTTTGTAATTCCTCCGAATGGACTGTGTG 1386  
DB 1659 AGAGATGCCATCTTCCCATCCCTGTGGCCTGTGATGCCCATCCCGAAAGACTGTGTG 1718  
QY 1387 CTGACGAGTGGAGGAGTGTGTCTGTTCCAGTCTCTTCAATTAATACTCAGAT 1446  
DB 1719 CTCAGCAGATGGTCTAGTGTCTCTCTGCTCACAACCTGCTCAGGGAACACAGAA 1778  
QY 1447 GGGAAACAGACAGCTCAAGAACTATCTCTGCG---ACTGGCTGGGGAAGGTGGAAGCCA 1503  
DB 1779 GGGAAACAGATACAGACAGATCCATTCTGSCCTATGCGGTGAAGAGGTGGAATTCG 1838  
QY 1504 TGTCCCTCTAGTACGCTCTCCAAAGAGCATCGTTTGTGTAATGACATTCCTGTATGAG 1563  
DB 1839 TGTCCAAATAGCAGTGTCTTGAAGAGAGTACGAAGCTGTATGAGCATCTCTGACAGTG 1898  
QY 1564 CTTCACTGGGAGACATCGCCTTGGGCGCTTGTCTGAGGACACATTTGTAACCTGCCCTT 1623  
DB 1899 TACCACTGGCAAACTGGTCCCTGGGCGCAGTGCAATGAGGACACCTCAGTATCGTCTTC 1958  
QY 1624 AATGCAACATTTGGCTGGAATGGAAGACCACTGTGTGTGATGAGCATTGAGCTCGGAGA 1683  
DB 1959 AACCAACTACGACTTGGAAATGGGAGGCGCTCTCTGCTGTGCGCATGCAAGCAAGAAA 2018  
QY 1684 GTCTTCTGTGCAAGAGTCACTGCGGCAAGTATGACCAAAAGATGTCCAGATTTCTACT 1743  
DB 2019 GTCATCTGTGTGAGTCAATGTGGCCCAAGTGGGACCCAAATAATGTCTCTGAAGCCTT 2078  
QY 1744 CGACTGAAACTGTGCGCCCTGTTTCTCCCATGCAAAAAGACTGTATTTGTCAGTGTCT 1803  
DB 2079 CGACTGAAACTGTGTAAGGCGCTTGTCTGCTTCTTGTGAAGAGACTGTATTTGTGACCCCA 2138  
QY 1804 TTCACTGAGTGGACACCTGCGCAAGGATGTGCCAAGCAGGAATGCCACAGT---AAAA 1860  
DB 2139 TATAGTGAATGAGCATCATGCCCTCTCTGCTGTGAAGAGGAGTCTCCAGTATCAGGAAG 2198  
QY 1861 CAGTCTCATACAGAACTCATCATCAAGAAAGACCCCAATGAGGCGCAGGAGTGCACAGAT 1920  
DB 2199 CAGTCTAGGCATCGGTCATCATTCAGTCTGCCACCAACCGGGGCGGAGACTGSCACAGAT 2258

1921	ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTCTGTGTCTCTATATCGTGGGAAG	1988
1922		
2259	CCCTCTATGAAGAGAGCCCTGTGAGGCACTTCAAGCGTGCCTCAAGACTACAGGTGGAAG	2318
1981	CCACAGAAATGGAGCCCTTGCATCTTAGTGCCAGAGTCTGTCTGGCAGGGAATAACGGGC	2040
2319	ACTCAAAATGGCGCAGATGTCAAATTAGTCCCTTGGAGCGTGCAACAAGACAGCCCTGGA	2378
2041	AGCAGTGAAGCCTGTGGAAAGGGTTTACAAACAAGAGCTGTCTCATGCATCTCTGATGAC	2100
2379	GCACAGGAAGGCTGTGGGCGTGGCGACAGGCAAGAGCCATTACTTGTGCAAGCAAGAT	2438
2101	AACCGGTACAGCAAAATGATGGAATCCCTCAAGCAGACAAAACGCACTGCCTCTCTCTGTG	2160
2439	GGAGGACAGCTTGAATCCATGAGTCCCTACAGTATGCAGGCCCTGTGCCAGCCCTTACC	2498
2161	CAAGAAATGCACAGTCCCATGTGCGAGAGACTGCACCTTCACTGCTTGGTCCAAAGTTTACG	2220
2499	CAGGCCCTGCCAGATCCCTTGCAGGATGACTGTCAATTGACCAAGCTGGTCCAAGTTTCT	2558
2221	CCCTGCTCCACGAACGTGTGAAGCCACAAAAGTAGGCGGCGACAGCTCAAGGAAAAAGC	2280
2559	TCATGCAATGGAGACTGTGTGTGAGTGGACAGCAAAAGGCACTCTTGTGGAATAAGT	2618
2281	AGAAAGAGGAGAAATGCCAGGATTCAGACTTTACCTCTAGTGGAGACAGAACTATGT	2340
2619	AAAAAGAGGAAAAATGTAAAAATTCCTATTTGTATCCCTGATTGAGACTCAGTATTGT	2678
2341	CTTGTGATGAATTTATATCCCAACCTTATGGAACCTGGTTCAGATTGCATTTCTCCAGAA	2400
2679	CCTTGTGACAAATAAATGCAACACTGTGGGAACTGTCAGACTGTATTTTACCAGAG	2738
2401	GGCAGAGGGAGCCTCACCGAGACTGCGGGTACAAGCAGACAGCAAAAGAAATGTGGAAA	2460
2739	GGAAAGAGTGGAAAGTGTGTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAA	2798
2461	GGCTGCGCTTTTCGAGCAGTAGCTGTCTGTATAAAATGGAAGACCTGTTCGACCCCTCC	2520
2799	GGATATCGTTTACCAAGCAATGGCATGCTAGACTCAAAATGGCAGGCTTGTGGAAAATCT	2858
2521	TTCTGCAGCAGCTCTGTTTACATTCAGAAAAATGTGTCAATTCCTGCCCATTTGATTGC	2580
2859	AGATGTAAACAGCCATGGTTTACATTTGAGGAGGCGTGCATCATCCCTGCCCTCAGACTGC	2918
2581	AAAGTTAAGCGATGGTCTAGTTGGGGTCTCGAGTTCACTTGTGGATTTGGAGTGAGA	2640
2919	AAAGTCAGTGTGTTCAACTGTGTCGCGCTGCAGCAAGTCTGTGGGAGTGGTGGAAG	2978
2641	ATTCGATCCAAATGGCTAAAAGAAAAACCTTACAATGGAGGACGACCATGTGCCAACTG	2700
2979	GTTCTGTTCTAAATGGCTGGTGA AAAACCATATAATGGAGGAAGGCTTGTGCCCAAACGT	3038
2701	GATC-----TCAAGBAATCAGGTACATGAGGCAAGTCCCATGTACAGTGAAGTGAATCAG	2754
3039	GACCATGTCAAACAGGACAGGTGTATGAGTTGTCCCATGTCCACAGTGCATGCAACCAAG	3098
2755	TATTCCTGGGTTGTAGAACACTGGTCTTCATGCAAAATCAAAATGAGCTGAGTCCCTG	2814
3099	TACCTATGGGTACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGTAATATCGGGAG	3158
2815	CGCTGTGGAGGAGGAACAATCTAGGAAAAATCAGATGTGT---GAATACTGGGATGT	2871
3159	AACTGTGGAGAGGGCGTGC AAAACCCGAAAAGTGAATGATGCAGAAATCAGCAGATGAC	3218
2872	GAAAGTGGAGCAGTGGATAGCAACCTGTGCAACCAAGATGAAATTCCTCCAGAAAACCCAG	2931
3219	CCTTCTGAACATGTAGAGGATTAACCTCTGTGACCCAGAGAGATGCCCCCTGGGCTTAGA	3278
2932	TCCTGTTCTTTATGTGTCCCAATAGATGTGTCACTGTGAGTGGGACATTTGGAGCAAA	2991
3279	GTGTGCAAAATTACCATGCCCTCAGGACTGTGTGATCTGATATCTGAAATGGGTCATGAGCCAA	3338

Qy	2992	TGCCACAGTCATGCGATCCCCACACAAT---GCAGAGAAAGAACTCGCCACACTGCTAAAGA	3048
Db	3339	TGTGTATTGCGTTTGCAAATCAAAGCAGTTTCGGCAAAAGGTCAGCTGATCCCATCAGACAA	3398
Qy	3049	CCATCACTGAACTCAAAGGACTTGTGTGTGAAGACTCAAGGTGCAGCGCTTCCCTCCCTGAAT	3108
Db	3399	CCAGCTGATGAAGGAGATCTTCGCCCTAATGCTGTTGAGNAAGAAACCTGTAACTCTGAAC	3458
Qy	3109	GAAAAATTGCTTCCAGTTCCAGTACAATCTTAACAAGAGTGGAGCAATGCCAGCTGAGTGAA	3168
Db	3459	AAAAACTGCTACCACATATGATTAATATGTAACAGACTGGAGTAGTACATGTACGCTGAGTGAG	3518
Qy	3169	AAGCACCCTGTGTCTCAAGSCGTCAGACCGCGCTCTCTAAGCTGTGTGTGCAGTGATGGC	3228
Db	3519	AAGCGATTGTGTGAAATGGAATATAAAAACAGGAATGTGGAATTTGTGTTCCGAAGTGATGGC	3578
Qy	3229	AAGCCAGTCAGCATGAGCAACCAATGTGACAGCAATAATTTGGAGAAGCGCCCAAGAGAAATGAGC	3288
Db	3579	AAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAAGAACTGGCCAGATGAAC	3638
Qy	3289	ATTCCTCGCTTGTGTGAATGCGTGTCTCACTGTTCAGCTCTTCAGGTTGGACGGCTTGGACA	3348
Db	3639	ACGTCTGCATGTGTGAATGCCCTGTGAACTGTTCAGCTTTCGATTGGTCTCCTTGGTCA	3698
Qy	3349	GAGTGTGTTCAGACACCTGTGGCCATGGAGGTTCGAATGAGCGCGACTCGAATTTATCATTTATG	3408
Db	3699	GAATGTCTCAAAACATGTGSCCTCACAGGAAATAATGATCCGAAGACGAACAGTGAACCCAG	3758
Qy	3409	CCAAACCAAGGAAGACGGCCATGCCCAACAGAGCTTACCCAGAGAGAAAACCTGCCCCA	3468
Db	3759	CCCTTTCAAGGTGATGGAAAGACCATCCCTTCCCTGATGGACCAAGTCCAAACCCCTGCCCA	3818
Qy	3469	GTGACCCCTCTGTCACAGCTGGGTCTCTGGCACTGTCTGCATGTAAATTTGGAGGGTGGGA	3528
Db	3819	GTGAAGCCTTGTTATCGTGGGCAATATGGCCAGTGTCTCCATGCCAGTGCAGAGGGCC	3878
Qy	3529	GACTGTGGGAAGGAGTTCCAGATCCGACGCTTTCTGTGATGGTCCACAGTGGTTCAATA	3588
Db	3879	CAGTGTGGGAAGGGACACAGAAACAAGGAAACATTTCTGTGTAGTAAGTGAATGGGTGAGCT	3938
Qy	3589	TCTCATGCAGCTGGACGTGTCCAGGATGCACTGTGTGGAGAAATGCCCCTT-----	3638
Db	3939	GATGATTTCHAGAAAGTGTGGATAGGAAATCTGTGCTGACATGGAATCAATATAGAT	3998
Qy	3639	-----TCAGACAGCATCCTGAAGCAGCTGTGTGTGTGTCCTTGCCTCCAGGAGACTGCCAT	3693
Db	3999	GGTAATAAAAAATATGTTCTGGAGGAATCCTGCAGCCAGCTTGGCCAGGTGACTGTAT	4058
Qy	3694	TTAAACAGATGTTCAGATGGAGCACATGTGAATTAACCTGCATGATGGAAGAACCTTT	3753
Db	4059	TTGAAGGACTGTGTTCTCCGTGGAGCGCTGTGTGAGCTGACCTGTGTGAATGGTGAGGATCTA	4118
Qy	3754	GAGACTGTGGGCGCCAGTCTAGATCAAGGACTTTTATAATTCAGTCTTTTGAGAACCCAA	3813
Db	4119	GGCTTTGGTGGAAATACAGTCCAGACCGGTGATTTATACAGAACTAGAGANTCAG	4178
Qy	3814	GACAGCTGCCCCCAAACAGGTTCTAGAAAACACGCCCTTGTACAGGAGGCAAAATGTTATCAC	3873
Db	4179	CATCTGTGCCAGAGCAGATGTTAGAAAACAAAATCATGTTATGATGGACAGTGTCTATGAA	4238
Qy	3874	TACACATGAAGAACAGTCTTTGGAAACAATAACGAACGAACCTGTAATGTTGCCAGCGTTCA	3933
Db	4239	TATTAATGGAATGCCAAGTCTCTGAAAAGGGCTCTTCCGAAACAGTGTGGTGTCAAAGGTCA	4298
Qy	3934	GATGGCGTTAATGTACAGGAGGCTGCTCCCTCAGGCCCGTCTCTGCTGCCATTCGGCAG	3993
Db	4299	GATGGTATAAAATGAACAGGGGGCTGCTTGGTGTATGAGCCAGCTGATGCCGACAGTCT	4358
Qy	3994	TGCATTCAGCGCTGCAGAAAACCTTTCTCTACTGTACACAGGGTGGAGTCTGTGGTTGT	4053
Db	4359	TGTAAACCCACCGTGTAGTCAACCCCACTGCTACTGTAGCGAGACAAAACATGTCCAATGT	4418
Qy	4054	GAGAAGGGCTATACAGAGATAATGAAATCAAATGGTTTCTCGATTACTGCA-----	4105

Db 4419 GAAGAGGGTACACTGAAGTCATGCTCTTCTTAACAGCACCCCTTGAGCAATGCACACTTATC 4478  
QY 4106 -----TGAAGTACCAGGCTCAGAGGATAAAAAAGCTGATGTGAAAAACCTTTCTG-G 4157  
Db 4479 CCGGTGGTGTATTTACCCACCATGGAGGACAAAGAGGAGATGTGAACCACTGGGCT 4538  
QY 4158 GAAAAACAGACCTGTGAAATTCAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
Db 4539 GTACATCCAAACCCCTCCAGTAAACCCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4598  
QY 4216 CCATTGATCCAGATGCGGAGTAAATTTGGTTTATGCGCTTTTCAGTGGCGCTTTT 4275  
Db 4599 CCATTGGGCGAGATGGAGACTTAAGACTGGTTTACGGTGTAGCAGCTGGGGCATTT 4658  
QY 4276 CTCATCATGATTTTCCPAATTTACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4332  
Db 4659 GTGTACTCATCTTTATTTGCTCCATGATTTATCTAGCTTGCATAAAGCCAAAGAACCC 4718  
QY 4333 CATCAAGCACACCTCCCGACAGAGCCTCTGACCTTAGCCTACGATGAGACTTAGAC 4392  
Db 4719 CAAGAAGGCAAAACACCCGACTGAAACCTTTAACCTTAGCTATGATGAGATGCCGAC 4778  
QY 4393 ATGTAA 4398  
Db 4779 ATGTAA 4784

RESULT 4  
AAD38694  
ID AAD38694 standard; cDNA; 6317 BP.  
XX AC AAD38694;  
XX AC AAD38694;  
DT 23-SEP-2002 (first entry)  
XX Human LP217 secreted protein encoding cDNA.  
XX Human; secreted protein; atherosclerosis; Alzheimer's disease; LP217;  
KW diabetic retinopathy; severe combined immunodeficiency; pancreatitis;  
KW rheumatoid arthritis; colorectal adenoma; haemolytic anaemia; cancer;  
KW reperfusion injury; arteriosclerosis; wound healing; transgenic animal;  
KW gene therapy; neoplasm; transgenic; psoriasis; ischaemia; carcinoma;  
KW chromosome /p21-p22; gene; ss.  
XX Homo sapiens.

Key Location/Qualifiers  
CDS 284..5158  
FT /\*tag= a  
FT /product= "Human LP217 secreted protein"  
FT sig\_peptide 284..328  
FT /\*tag= b  
FT mat\_peptide 329..5155  
FT /\*tag= c  
FT /product= "Mature human LP217 secreted protein"

XX WO200226801-A2.  
XX 04-APR-2002.  
XX 14-SEP-2001; 2001WO-US026026.  
XX 28-SEP-2000; 2000US-0236088P.  
XX (ELIL ) LILLY & CO ELI.  
XX Su EW, Wang H;  
XX WPI; 2002-471259/50.  
XX P-ESDB; AAE23579.  
XX Novel proteins and polynucleotides of secreted proteins useful for

PT treating various diseases e.g. rheumatoid arthritis, cancer, psoriasis,  
PT diabetic retinopathy, arteriosclerosis, ischemia or reperfusion injury.  
XX Claim 1; Page 109-117; 145pp; English.  
XX The invention relates to human secreted polypeptides designated LP095,  
CC LP191, LP217, LP220, LP221, LP222, LP229, LP237 or LP238 and nucleic acid  
CC molecules encoding such polypeptides. Novel secreted proteins of the  
CC invention are used for treating diseases such as atherosclerosis,  
CC Alzheimer's disease, diabetic retinopathy, psoriasis, pancreatitis,  
CC arteriosclerosis, rheumatoid arthritis, colorectal adenomas, severe  
CC combined immunodeficiency, ischaemia, carcinomas, haemolytic anaemia,  
CC reperfusion injury, neoplasms and cancer especially liver cancer. They  
CC are also used for wound healing. Polynucleotides of the invention can be  
CC used to generate transgenic animals or knock out animals, which in turn,  
CC are useful in the development and screening of therapeutically useful  
CC reagents for use in the treatment of diseases associated with LP  
CC polypeptide associated activity. They are also used in gene therapy. The  
CC present sequence is human LP217 secreted protein encoding cDNA. LP217  
CC Gene is located on chromosome 7p21-p22  
XX Sequence 6317 BP; 1776 A; 1442 C; 1617 G; 1482 T; 0 U; 0 Other;

Query Match 28.0%; Score 1232.8; DB 6; Length 6317;  
Best Local Similarity 57.7%; Pred. No. 0;  
Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;  
QY 10 TGCATTGAGAGCTGAACCGAACTGTGGTTCCTCAATGAATATCGGAACACTTGGCCCTT 69  
Db 662 TGCATCCAGAAAGCAAGACATTCCTGCGGAGGATATCATCTGTGATCTTTGAGCCC 721  
QY 70 CAGCCTCTCAGAACAGGCTTGCTCATCTTCTGCTCCCGGATTTGTAGTATCTTGAG 129  
Db 722 AAGCTCTCTCGAGCAGGCTTGCTCATCTTCTGCGCAAGATTGCACTGCTCTGAA 781  
QY 130 TTCTTACCATCGTCCAACTGTAGCAAGGATGTGGGAAGAAATTCGAGATAGAACTGCG 189  
Db 782 TTTTCTGCTGCTCGAATGTCTCAAGACCTGCGGAGCGGGCTCCAGACCGGACGCGT 841  
QY 190 CGGGTCATAGCTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGACTCAAGA 249  
Db 842 CATGTGTGGCGCCCGCCAGTTCGAGGCTCTGGTGTCTCAACCTGACGAGTTCAG 901  
QY 250 GCTGTGATGTCCCATTTCTGCTCTTGGGGAAGAGGAATATACATTTAGCCTTAAG 309  
Db 902 GTGTG-----CCAATCCAGTCCATGCGAGCGGAGGCTCAGGTACAGCCTGCAT 952  
QY 310 GTTGACCATGAGTAAATGCAGACTGCTCATCTTAAAGAAATTAATCCAGCGGAAGA 369  
Db 953 GTGGGCGCTTGAGCACCCTGCTCAATGCCCCACTCCGCAAGTAAGACAGCAAGAGA 1012  
QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTTCAATGAG 402  
Db 1013 CGCGGAAGATAAAGAACGGGAAAGACCGCAGCAAAAGGAGTAAAGGATCCAGAGCC 1072  
QY 403 CGAGTCACCTTTAAACATCAAGTTAG-----AAGCATCATCATCTTCAAGTCT 453  
Db 1073 CGCGAGCTTTATTAAGAAAAGAGAAACAGAAACAGGCGAGAACAGCAAGAACAAATAT 1132  
QY 454 TGGCAATAGAGATAGGTTATCAAAACCGGAGGTTTCTGTACAAGAAAGTGTGGAACA 513  
Db 1133 TGGGACATCCAGATTTGATATCAGACAGAGGTTATGTGATTAACAAGCGGGGAA 1192  
QY 514 AATGCTATGTTAAGCCTTTGCTTCAAGATTTCTTCCCATGAGTGTTCAGTCTGCTGATC 573  
Db 1193 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGAAGCTTCCAATGACCTTCCAGTCTGTGTG 1252  
QY 574 ATGCCCAAGAGCTGTGAACCTCCAGTGTGCTCTCTGAGCCCTGCTCAAGACATGC 633  
Db 1253 ATCAACAAAGAGTGCAGGTTTCCAGTGTGTCAGATGAGGAGCCCTGCTCAAAACATGC 1312  
QY 634 CTTTCAGGAGTCTCTTCCAGGATTTTAGGAGCAGGAGCCGGAACGTAAGCAATGGCT 693



1313 CATGACATGGTGTCCCTCGAGGCACTCGTGTAAAGACACGAACCAATCAGGCGAGTTTCCC 1372  
694 ATTGGAGGTGGAAGAGTGTCTCTGAACCTTCTTGAAAGAGCGCTGCATTTGTGAAGGA 753  
1373 ATTGGCAGTGAAGAGAGTGTCCAGAAATTTGAAGAAAGAACCCCTGTTGTCTCAAGGA 1432  
754 GAACTTCTGACGAATGTCCAGGTATCTCTGAGNACTTCTGAATGGAAGMATGCCAA 813  
1433 GATGGAGTTGTCCCTGTGCCACGTATGGCTGGAGAACTACAGATGGACTGAGTGGCGGT 1492  
814 GTCTCTCTCTCTCTCTGAGCAGCAGAGATCCCACTGGCATGTGACGGGACCCGCTGTGTGC 873  
1493 GTGGACCTTTGCTCAGTCAGCAGGACAGAGCGCGGCAACCAAGACGCCCTCTGTGGA 1552  
874 GTGGGATCAGACCCGGGAGGTGTACTGTGCCAG-----AGGTACAGCA 921  
1553 GGGGGCATCCAGACCCGAGAGGTGTACTGCTGTCAGGCCAAGCAAAACCTCTCTCAAA 1612  
922 GCTGCCGCACTGAGGCGCAGGAAGTCTCTAGACCTGTGGAAGAGGCAATTATGTGTGGA 981  
1613 TTAAGTACCAACAGAACAAAGAGGCTCAAGCCMATGGACTTAAATTTATGCACTGGA 1672  
982 CCGGCCCCCTCTCTCAGCTCTGCAATATCCCTTGCTCTACGGACTGCATAGTATCT 1041  
1673 CCTATCCCTAATACTACACAGCTGTGCCACATTCCTTGTCCAACCTGAATGTGAAGTTCA 1732  
1042 TCTGTGTGAGCTGGGCTGTGCATCCATGAAGAACTGTCAACCTCAGGCGGAAGAAA 1101  
1733 CTTGTGTGAGCTGGGACCTTGTACTTATGAAGAACTGTGAATGATCAGCAAGGGAAGAAA 1792  
1102 GGAATTTAGAACAGGCGAGCG-----CCATGTCCTCATGGAATCTACAGGCGCTGCA 1152  
1793 GGCTTCAAACTGAGGAAGCGCGCATTAACATGAGCCCACTGGAGGCTCTGGGGTAACC 1852  
1153 GGGCATTTGCCCTCATTTGTGTGAGTGTCTTCTGTGAGGATCCAAATGTGTACCGATGG 1212  
1853 GGAATCTGCCCTCACTTACTGTGAAGCCATTCCTGTGAAGACCTGCTGTATGACTGG 1912  
1213 ---CTGGCATCAGAAGGATCTGTTTCCCTGTATCATGGA---AATGTGGCTGGGACAT 1266  
1913 AAGCGGTGAGACTGGGACCTGCGAGCCAGATAACGGAAGGAGTGTGTCTCAGGACG 1972  
1267 CGTATTTGGAAGCGCTGTGCGAATGACCGGGAAGATGTATCAGGAGTCTTTGC 1326  
1973 CAAAGTTCAAGAGTGTGTGCAATCAACAGTGTATGGAAGAAGTTGACAGACAGCTGTGC 2032  
1327 CCAGTTCCCTCCTCTGAGAGGAGTCTTGTGAAATTCCTGCGCAATGGAAGTGTGTG 1386  
2033 AGAGATGCCATTTCCCATCCCTGTGGCTGTGATGCCCATGCCCAAGAGACTGTGTG 2092  
1387 CTGAGCGAGTGACGAGTGTGTCTCTGTTCCAGTCTGTTCAAATTAAGAACTTCAGAT 1446  
2093 CTCAGCAGATGTCTACGTGTGTCTCTGCTCACACACTGCTCAGGGAAGAACGACAGAA 2152  
1447 GGGAAACAGACAGCTCAAGAACTATCTGGC---ACTGGCTGGGGAAGTGGGAAGCCA 1503  
2153 GGGAAACAGATACGAGCAGATCCATTTGGCTATGCGGGTGAAGAAGTGGAAATTCGC 2212  
1504 TGTCCCTCTAGTCAGGCTCTCCAAGACATCGTTTGTGTAAATGACCATTCCTGTATGCA 1563  
2213 TGTCCAAATAGCAGTGTCTTGAAGATGACGAAGCTGTATGAGCATCTTGCACAGTG 2272  
1564 CTTTCTGAGGACATTCGCTTGGGGCCCTTGTCTGAGGACACATTCGTAACCTGCGCTT 1623  
2273 TACCACTGSCAAACTGGTCCCTGGGGCCAGTGCATTTAGGACACCTCAGTATCGTCTTC 2332  
1624 AATGCAACCATTTGGTGGAAATGGAAGCCAGCTGTGTGTGAGGCAATTCAGACTCGGAGA 1683  
2333 AACAACAATTACACTTGGATGGGAGGCTCTCTGTCTGCGCATTCGACAGCAAGAAAA 2392  
1684 GTCTTCTGTGTCAGAGTCACTGCGGACAAGTAATGACCAAGATGTCCAGATTTCTACT 1743  
2393 GTCACTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAGAAATGTCTTGAAGGCTT 2452

1744 CGACCTGAACACTGTGGCCCTCTGTTTTTCTCCATGCAAAAAAGACTGTATTGTACTGCT 1803  
2453 CGACCTGAACACTGTGAAGCCTTGTGCTTCTTGTGAAGAGACTGTATTGTGACCCCA 2512  
1804 TTAGTGTGAGTGGAACCCCTGCCCAAGATGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
2513 TATAGTACTGSGACATCATGCCCTCTCTCGTGTAAAGAGGGGACTCCAGTATCAGGAAG 2572  
1861 CAGTCTCGATACAGAAATCATCATCCAAAGAGCAGCAATGGAGGCCAGGAATCCAGAT 1920  
2573 CAGTCTAGGATCGGTTCATCATTCAGCTGCCAGCCAAAGGGGCGAGACTGCAAGAT 2632  
1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTTCTTGTGTCTGTATATCCGTGGAAG 1980  
2633 CCCCTCTATGAAGAGAGGCTGTGAGGCACCTCAAGCGTGCACAAAGCTACAGGTGGAAG 2692  
1981 CCACAGAAATGSGAGCCCTTCATCTTAGTGCCAGAGTCTGTCTGTGGCAGGAAATAACGGGC 2040  
2693 ACTCAAAATGGCGAGATGCCAAATAGTCCCTTGGAGCGTGCAACAAGACAGCCCTGGA 2752  
2041 AGCAGTGAAGCCTTGTGAAAGGGTTTACAAACAAGAGCTGTCTCATGCATCTCTGTATGAC 2100  
2753 GCACAGGAAGGCTGTGGGCTGTGGCGACAGGCAAGAGCCATTACTTGTGCAAGCAAGAT 2812  
2101 AACGGTCACAGAAATGATGGAATGCTCAAGCAGACAAACGGCATGCTCTCTTTGTG 2160  
2813 GGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGCAAGGCCCTGTGCCAGCCCTTACC 2872  
2161 CAAGAATGCAACAGTCCCATGTGCGAAGAGCTGCACCTTCACCTTGTGGTCCAAAGTTTACG 2220  
2873 CAGGCTGCCAGATCCCTGCCAGGATGACTGTCAATTGACAGCTGTGTCRAAGTTTCT 2332  
2221 CCTGTCTCAGAACTGTGGAAGCCAAAGAGTAGCGGGGACAGCTCACAGGAAAGC 2280  
2933 TCATGCAATGAGACTGTGTGCACTTAGSACCAGAAAGCGCACTCTTGTGTGAAAAAGT 2992  
2281 AGAAAGAGGAGAAATCCAGAGGATCTGACCTTTTACCTCTAGTGAGACAGAACTATGT 2340  
2993 AAAAGAGGAAATGTAAGAAATTCATTTGTATCCCTGATTCAGACTCAGTATTGT 3052  
2341 CTTGTGATGAATTTATATCCCAACTTATGGAACCTGTGTAGATTCATTTCTTCCAGAA 2400  
3053 CTTGTGACAAATATATGACAACTGTGGGAACTGTGTGCACTGTATTTTACCAGAG 3112  
2401 GGCAGAGGGAGCCTCACCGAGACTGCGGTACAGCAGACAGCAAGAAATGTCGGAGAA 2460  
3113 GGAAGTGGAGTGTGTGGAATGAAGATCAAGGAGACATCAAGGAATCGCGACAA 3172  
2461 GGCCTCGCTTTTCAGCAGTAGCCTTTCTGATAAAATGGAAGACCTTTGACCCCTCC 2520  
3173 GGATATCGTTTACCAAGCAATGGCATCTACGATCAAAATGGCAGGCTTGTGAAACATCT 3232  
2521 TTCTGAGCAGCTCTGTTACATTCAGAAATGTGTCTATTCCTGCCCTGCTTGTGAG 2580  
3233 AGATGTAACAGCCATGGTTACATTTAGGAGGCGCTGCATCATCCCTGCCCCCTCAGACTGC 3292  
2581 AAGTTAAGCGAATTTGTGTAGTTGGGGCTTTGCGAGTTCTATTTGTGGAATTTGAGTGAGA 2640  
3293 AAGCTCAGTGAGTGTTCGAAGTGTGCGCTGCAGCAAGTCTCTGTGGAGTGTGTGAG 3352  
2641 ATTGCAATCMAATGGCTTAAAGAAACCTTCAATGAGGAGGAGCACTGTCTCCAAACTG 2700  
3353 GTTCTGTTAAATGGCTGTGAAAGCAATATAATGAGGAGGCGCTTTGCCCAAACTG 3412  
2701 GATC-----TCAAGAAATCAGGTACATGAGGAGTCCCATGTTTACAGTCAAGTCAATCAG 2754  
3413 GACCATGTCAACCGGACAGGTGTATGAGTGTGTCCTCCATGCCACAGTCACTGCAACAG 3472  
2755 TATTCCTGGTGTAGAACACTGGTCTTCTCATGCAAAATCAACAAATGAGCTGAGGTCCCTG 2814  
3473 TACCTATGGTCAACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTTGTGAATATGCGGAG 3532

```
QY 2815 CGTGTGGAGGAGAAACAATCTAGAAAATCAGATGTGT---GAATCTGGGATGGT 2871
Db 3533 AACTGTGGAGGGCGTGCAAAACCGAAAAGTGAGATGCATGCAGATAACAGCAGATGGC 3592
QY 2872 GAAGGTGGAGCAGTGGATAGCAACCTCTGCACACAGATGAAATTCCTCCAGAAACCCAG 2931
Db 3593 CCTTCTGAAATGTAGAGATTACTCTGTGACCCAGAGAGATGCCCTCTGGGCTCTAGA 3552
QY 2932 TCCGTGTTCTTTATGTGCCAATGAGTGTCTATGTCTGAGTGGGACTTTTGGAGCAAA 2991
Db 3653 GTGTGCAAAATFACCATGCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGACCAA 3712
QY 2992 TGCCACAGTCACTGGATCCCAACAAAT---GCAGAGAGAACTGCCACCTGCTAAGA 3048
Db 3713 TGTGTTTTCCTTTGCAATCAAGCAGTTTCGGCAAGGTCAGCTGATCCCATCAGACAA 3772
QY 3049 CCATCACTGAATCAAGGACTTTGTGCTGAAGACTCACAGGTGCAGCCTTGCCCTCCCTGAAT 3108
Db 3773 CCAGTGATGAAGGAGATCTTGCCCTAAATGCTGTTGAGAAAGAACCTGTAACTGAAC 3832
QY 3109 GAAATGTCTCCAGTCCAGTACAATCTAACAGAGTGGAGCACATGCCAGCTGAGTGAA 3168
Db 3833 AAAAACTGCTACCACTATGATTATTAATGTAAACAGACTGGAGTACATGTCACTGAGTGAG 3892
QY 3169 AACGCACCTGTGTCAAGGCTCAGAACCCGCTGCTAACTGTGTGTGAGTGAATGGC 3228
Db 3893 AAGCAGTTTGTGAAATGGAATTAATAACAGGATGTGGATTGTTCGAATGATGGC 3952
QY 3229 AAGCAGTCAAGTGGACCAATGTGAGCAGCAATTAATTGGAGAGCCCAAGAGATGAGC 3288
Db 3953 AAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAGAACTGSCAGATGAAC 4012
QY 3289 ATTCCTCTGTGTGGATGGTGTCAACTGTCACTCTCAGGTCGACGGCTTGGACA 3348
Db 4013 ACGTCTGCAATGGTGAATGCCCTGTGAATCTGAGCTTCTGATGCTCTCCTTGTGCA 4072
QY 3349 GAGTGTTCACAGACTGTGGCCATGGAGGTGCAATGAGCCGACTCGAATTTATCATATG 3408
Db 4073 GAATGTTCTCAACATGTGGCTCACAGGAAATATGATCCGAGCAACAGTGACCCAG 4132
QY 3409 CCAACCCAGGAGAGAGACGGCATGCCCAACAGAGCTTACCCAGAGAAACTGCCCA 3468
Db 4133 CCTTTCAAGGTGATGGAGAACCATGGCTTCCCTGATGACCACTTCCAAACCTTGCCCA 4192
QY 3469 GTGACCCCTCTCAGCTGCGGTCTTGGCAACTGGTCTGCATGTAATTTGGAGGTGGA 3528
Db 4193 GTGAAGCCTTGTATCGTGGCAATATGGCCAGTGGTCTCATGCCAAGTGCAGGAGGCC 4252
QY 3529 GACTGTGGGGAAGGAGTTACAGATCGGAGCCTTCTCGATGTCATGTCACAGTGGTTCAATA 3588
Db 4253 CAGTGTGGAGAAAGGACACAGAACAGGAACAATTTCTTGTGTAGTAAGTGAATGGGTCACT 4312
QY 3589 TCTCATGCACTGCGAGTGTGAGGATGCACTGTGTGAGAAATGCCCTT----- 3638
Db 4313 GATGATTTACGAAAGTGTGTGATGAGAAATCTGTGCTGACATTGAACCTATTATAGAT 4372
QY 3639 -----TCAGGACAGATCCTGAAGCAGCTGTGTCTGTGCTTCCAGAGACTGCCAT 3693
Db 4373 GGTAAATAAATATGTTTCTGGAGGAATCCTGSCAGCCAGCCTTCCCAAGTGAATTTAT 4432
QY 3694 TTAACAGAAATGTCAGAGTGCAGACATGTGCAATTAACCTGCAATTTGATGGAAGAGCTTT 3753
Db 4433 TTGAAGGACTGTTCTCTGGAGCTGTGTGAGTGAATGCTGCTGATGAGGATCTA 4492
QY 3754 GAGACTGTGGCCGCCAGTCTAGATCAAGGACTTTTATAATTCAGTCTTTTGGAGAACCAA 3813
Db 4493 GCGTTTGTGGAATACAGGTACAGATCCAGACCGGTGATTATACAGAACTAGAGAATCAG 4552
QY 3814 GACAGCTGCCCCACAGGTTCTAGAAACACGCCCTTGTACAGAGGCAATGTTATCAC 3873
Db 4553 CATCTGTGCCCAGACGACATGTTAGAAACAAAATCATGTTATGATGACAGTGTCTATGAA 4612
QY 3874 TACACATGAAAGCAAGTCTTTTGGAAACAATAACGAACGAACTGTATGTTGTCACGCTTCA 3933
```

```
Db 4613 TATAATGATGGCCAGTGTCTTGGAAAGGGCTCTTCCCGAACAGTGTGTGTCAAAAGTCA 4672
QY 3934 GATGCGGTAAATGTACAGAGGAGTGTCTCCCTCAGGCCCGTCTCTGTGCTGCATTCCGAG 3993
Db 4673 GATGATTAATGTAAACAGGGGCTGTGTGTGATGAGCAGCCTGATGCCAGAGTCT 4732
QY 3994 TGCATTCCAGCTGCAGAAAACCTTTTCTCTACTGTACACAGGTGGAGTGTGTGTTGT 4053
Db 4733 TGTAAACCCACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAACATGCCATTGT 4792
QY 4054 GAGAGGGCTATACAGAGATTAATGAATCAATGGTTTCTGGATTACTGCA----- 4105
Db 4793 GAAGAAGGTACACTGAAGTATGTTCTTAAACAGACCTTGAACATGCACACTTATC 4852
QY 4106 -----TGAAGTACCAGGCTCAGAGATAAAAAGCTGATGTGAATAAACTTTCTG-G 4157
Db 4853 CCGGTGGTGTATTATCCCACTGAGGACAAAAGAGAGATGTGAATAACCACTCGGCT 4912
QY 4158 GAAACACAGACCTGTGAATTCAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215
Db 4913 GTACATCAACCAACCCCTCCAGTAAACCCAGAGACCGGGAAGACCTGGTTTCTACAG 4972
QY 4216 CCACCTTGATCCAGATGGCCGAGTAAAAATTTGGGTTTATGGCGTTTCAGGTGGGCTTTT 4275
Db 4973 CCATTTGGCCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGCAATT 5032
QY 4276 CTCATCATGATTTTCTTAATTTACTTCTTACCTTGTGTTTTCGAAGAGCCA---AAACCA 4332
Db 5033 GTGTTACTCATCTTTTATTTGTCTCCATGATTTATCTAGCTTGCNAAGGCCAAAGAACCC 5092
QY 4333 CATCAAGCACACCTCCCAACAGAAAGCTCTGACCTTAGCTACGATGGAGACTTAGAC 4392
Db 5093 CAAAGAGCAAAACAAACCGACTGAACCTTTAACCTTAGCTATGAGAGATGCCGAC 5152
QY 4393 ATGTAA 4398
Db 5153 ATGTAA 5158

RESULT 5
AAF30188
ID AAF30188 standard; cDNA; 6373 BP.
XX
AAF30188;
AC
XX
XX
DT 30-APR-2001 (first entry)
XX
DE Clone 21433858 encoding secreted protein SECPI.
KW SECPI; secreted protein; human; diagnosis; gene therapy; lung cancer; ss.
XX
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 235..5001
FT /*tag= a
FT sig_peptide 235..303
FT /*tag= b
FT mat_peptide 304..4998
FT /*tag= c
XX
XX WO200105971-A2.
XX
XX 25-JAN-2001.
XX
XX 20-JUL-2000; 2000WO-US019890.
XX
XX 20-JUL-1999; 99US-0144722P.
XX 29-NOV-1999; 99US-0167785P.
XX 19-JUL-2000; 2000US-00619252.
XX
XX (CURA-) CURAGEN CORP.
PA
```



Db 2116 TACCACTGGCAAACTGGTCCCTGGGGCCAGTGCAATTGAGGACACCTCAGTATCGTCCTTC 2175  
QY 1624 AATGCAACCAATTTGGCTGGAATGGAGAGACCCACGCTGGTGTAGGCATTCAGACTCCGAGA 1683  
Db 2176 AACCAACTACCACTTGGATGGAGAGCCCTCCCTCTCTGTGGCATGCGAGCAAGAAA 2235  
QY 1684 GTCTTCTGTGTCAAGATGTCAGTGGGCAAGATTAATGACCAAAAAGATGCCAGATTTCTACT 1743  
Db 2236 GTCACTCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCCAAAAATGTCTGAAAGCCCT 2295  
QY 1744 CGACCTGAACTGCGCCCTGTTTCTCCCATGCAAAAAGCTGTATGTGACTGCT 1803  
Db 2296 CGACCTGAACTGTAGGCCCTGTCTCTCTTGTGAAGAGGACTGTATGTGACCCCA 2355  
QY 1804 TTCACTGAGTGACACCCCTGCCAAGGATGTGCCAAGCAGCAAAATGCCACAGT---AAAA 1860  
Db 2356 TATAGTGACTGACATCATGCCCTCTCTCGTGTAAAGAGGGACTCCAGTATCAGGAAG 2415  
QY 1861 CAGTCTCGATACAGAAATCATCATCAAGAGCAGCCAAATGGAGCCAGGAATGCCCATGAT 1920  
Db 2416 CAGTCTAGGCATCGGGTCACTCAITCAGCTGCCAGCCAAACGGGGGCCGAGACTGCACAGAT 2475  
QY 1921 ACCTTATATGAGGAGAGAGAGTGTGAAGATGTTCCTTGTGTCTGTATATCGTGGAG 1980  
Db 2476 CCCCTCTATGAAGAGAGGCCCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGTGGAG 2535  
QY 1981 CCACAGAAATGAGCCCTTGATCTTATGTGCGAGAGTCTGTGTGGCAGGAAATACGGGC 2040  
Db 2536 ACTCACAAATGGCCAGATGCCAATTAAGTCCCTTTGGAGCGGTGCAACAAAGACAGCCCTTGA 2595  
QY 2041 AGCAGTGAAGCCTGTGAAAGGGGTAAACAAAGAGCTGTCTCATGCTCTGTATGATGAC 2100  
Db 2596 GCACAGGAGCTGTGGCCCTGGCGCAGCAGCAGAGCCATTAAGTGTGCGAAGCAGAT 2655  
QY 2101 AACCGGTACAGAAATGATGAATGCTCAAGCAGACAAACGSCATGCTCTCTTTGTG 2160  
Db 2656 GGAGGACAGGCTGAATCCATGAGTGCTACAGTATGACGGCCCTGTGCGAGCCCTTACC 2715  
QY 2161 CAAGAAATGCACAGTCCCATGTCGAGAGAGCTGCACCTTCACTGCTGGTCCAGTTTACG 2220  
Db 2716 CAGGCTGCGAGATCCCTGCGCAGATGACTGTCAATGACAGCTGGTCCAGTTTCT 2775  
QY 2221 CCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGTGGCGGCGACAGCTCAAGGAGAAAAGC 2280  
Db 2776 TCATGCAATGGAGACTGTGTGTCAGTTAGGACCAAGAGCGCACTCTTTGTTGAAAAAGT 2835  
QY 2281 AGAAGAGGAGAAATGCCAGATTCAGACCTTTACCTCTAGTGGAGACAGACTATGT 2340  
Db 2836 AAAAGAGGAGAAATGTAAAAATTCACATTTGTATCCCTGTATGAGACTCAGTATTGT 2895  
QY 2341 CCTTGTGATGAATTTATATCCCAACCTTATGGAACCTGGTCAGATTGCATTTCTCCAGAA 2400  
Db 2896 CCTTGTGACAAATATATGACACACCTGTGGGAACTGGTCAGACTGTATTTTACCAGAG 2955  
QY 2401 GGCAGAGGAGCCCTCACCGAGGACTGGGGGTACAAGCAGACAGCAAAAGATGTGAGAA 2460  
Db 2956 GGAAGAGTGAAGTGTGCTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGCAAA 3015  
QY 2461 GGCCTGCGCTTTTCGAGCAGTAGCTGTCTGTATAAAAATGAAGACCTGTGACCCCTCC 2520  
Db 3016 GGATATCGTTACCAAGCATGGCATGCTACATCAAAATGGCAGCTGTGTGAAACATCT 3075  
QY 2521 TTCTGACAGCTCTGTTACATTAAGAAAAATGTGTCAATCCCTGCCCATTTGATTCG 2580  
Db 3076 AGATGTAAACAGCCATGGTTACATTTGAGAGGCCCTGCATCATCCCTGCCCTCAGACTGC 3135  
QY 2581 AAGTTAAGCAGTCTGCTAGTTGGGCTCTTGCAGTTTCATCTGTGAAATGGAGTGA 2640  
Db 3136 AAGCTCAGTGAAGTCTCAACTGTGCGCTGCGAGCAAGTCCCTGTGGAGTGGTGAAG 3195  
QY 2641 ATTCGATCCAAATGGCTAAAAAGAAAAACCTTTACAATGGAGGACACCATGTCCCAAACTG 2700

Db 3196 GTTCGTTCTAAATGGCTCGTGAAAAAACCATATAATGAGGAGAGCCCTTGCCCAAACTG 3255  
QY 2701 GATC-----TCAGAAATCAGSTACATGAGCAGTCCCATGTTACAGTGTAGTGAATCAG 2754  
Db 3256 GACCATGTCAACAGGCAAGGTGTATGAGTTGTCCCATGCCAGTGTAGTGCACACCAG 3315  
QY 2755 TATTCCTCGGTTGTAGAACACTGGTCTTCAATGCAAAATCAACAATGAGCTGAGTCCCTG 2814  
Db 3316 TACCTATGGGTACACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAG 3375  
QY 2815 CGCTGTGAGGAGGAGAACCAATCTAGGAAATCAGATGTGT---GAATACTGCGGATGGT 2871  
Db 3376 AACTGTGAGAGGGCGTGCAAAACCCGAAAGTGTGATGCAATGCAAAATACAGCAGATGGC 3435  
QY 2872 GAAGTGGAGCAGTGGATAGCAACCTGTGCAACAGGATGAAATTCGCCCAAGAAACCCAG 2931  
Db 3436 CCTTCTGAACATGTAGAGGATTTACCTCTGTGACCCAGAGAGATGCCCTGGCTCTAGA 3495  
QY 2932 TCCTGTTCTTATGTGTGCCAATGAGTGTGTCTGATCTGAGTGGGACCTTTGGAGCAAA 2991  
Db 3496 GTGTGCAAAATACCATGCCCTTGAGACTGTGTGTATCTGAAATGGGTCCATGGAACCCAA 3555  
QY 2992 TGCCCAACAGTCAATGCGATCCCCACACAAT---GCAGAGAGAACTCGCCACCTGTCTAGA 3048  
Db 3556 TGTGTTTTCCTTGCATCAAGCAGTTTCGGCAAGGTGAGTGTGATCCCATCAGACAA 3615  
QY 3049 CCATCACTGAATCAAGACTGTGTCTGAAGACTCAAGGTGACAGCTTGCCCTCTCTGAAT 3108  
Db 3616 CCAGCTGATGAAGGAAGATCTTGCCCTAATGCTGTGAGAAAGAACCTGTAACTGAAC 3675  
QY 3109 GAAATTCCTCCAGTTCAGTGTCAATCTAACAGATGGACACATGCCAGCTCAGTGA 3168  
Db 3676 AAAAATCTCTACCACTATGATTAATGTAAACAGCTGGAGTACATGTACCTGAGTNG 3735  
QY 3169 AACGACCCCTGTGTCAAGCGCTCAGGACCCGCTGTCAAGCTGTGTGTGAGTGTGAGTGC 3228  
Db 3736 AAGGAGTGTGTGGAATGGAATAAACAAGGATGTGGATTTGTGTGGAAGTGTAGTGC 3795  
QY 3229 AAGCAGTCAGACTGAGCAATGTGAGCAGCATAAATTTGGAGAGCCCCAGAGATGAGC 3288  
Db 3796 AAGTCAGTTGACTGAAATATTTGGAAGCGCTTGGCTTGGAGAGAACTGGCAGATGAAC 3855  
QY 3289 APTCCCTCTTGTGTGAATGGTGTCAACTGTCTCAGCTCTCAGGGTGGACCGCTTGGACA 3348  
Db 3856 AGTCTCTGATGTGGAATGCCGTGAACTGTGAGCTTTCTGATTGTCTCTCTTGGTCA 3915  
QY 3349 GAGTGTTCACAGACCTGTGGCATGGAGTGCATGAGCCGAGTCTGATTTATCATATG 3408  
Db 3916 GAATGTTCTCAACATGTGGCTCACAGGAAAAATGATCCGAAAGCAGTGAACCCAG 3975  
QY 3409 CCAACCCAGAGAGAGGAGCGGCATGCCCAACAGAGCTTTACCCAGGAGAAACCTGGCCA 3468  
Db 3976 CCTTTCAGGTGTGGAAGACCATGCCCTTCCCTGATGGACAGTCCAAACCCCTGCCCA 4035  
QY 3469 GTGACCCCTCTGTACAGCTGGGTCTTGGCAACTGTCTGATGTAAATTTGGAGGTGGA 3528  
Db 4036 GTGAAGCCCTTGTATTCGGTGGCAATATGSCAGTGTCTCCATGCAAGTGCAGAGGCC 4095  
QY 3529 GACTGTGGGAGGAGTTTCAGATCCGAGCCTTTCTCTGATGGTCCACAGTGGTTCAATA 3588  
Db 4096 CAGTGTGAGAAAGGAGCCAGAACAGGAACATTTCTGTGTAGTGTAGTGTGGTCACT 4155  
QY 3589 TCTCATGAGCTGGAGTGTGTGAGGATGCACTGTGTGGAGAAATGCCCTT----- 3638  
Db 4156 GATGATTTTCAGCAAGTGTGTGATGAGGAATTTCTGTCTGACATTTGAATCATTTATAGAT 4215  
QY 3639 -----TCAGGACAGCATCTGAAGCAGCTGTCTGTGCTTCCAGGAGAGTGCCT 3693  
Db 4216 GGTAAATAAATAATGGTCTCTGGAGGAATCTTCGAGCCAGCCCTTGCCCAAGTGTAT 4275  
QY 3694 TTAACAGAAATGTCAGAGTGGAGCAGATGTGAATTAACCTCATTTGATGGAGAGCTTT 3753  
Db 4276 TTGAAGGACTGTCTTCTCTGGAGCCTGTGTGAGCTGACCTGTGTGAATGGTGAAGATCTA 4335

3754 GAGACTGTGGCGCCGCTAGATCAAGACTTTTATTAATTCAGTCTTTTGGAGAACCA 3813  
 4336 GGCTTTGGTGGAAATACAGGTTCAGATCCAGACCGGTGATTATATCAAGAACTAGAGAACTCAG 4395  
 3814 GACAGCTGCCGCCCAACAGAGGTTCCTAGAAACACAGCCCTTGTACAGGAGGCAAAATGTTATCAC 3873  
 4396 CATCTGTGCCAGACAGATGTTAGAAACAAATCAATCAATGATGAGACAGTGTCTATGAA 4455  
 3874 TACACATGGAAACAGACTTTTGGAAACAATAACAAACAGAACTGTATGTCGCCAGGCTTCA 3933  
 4456 TATTAATGGATGCCAGTGTCTTGGAAAGGCTCTTCCGGAACAGTGTGTCAAAGGTCA 4515  
 3934 GATGGCGTTAATCTCACAGAGGCTGCTCCCTCAGCCCGCTCTCTGCTCCATTTGGCAG 3993  
 4516 GATGGTATAATGTAACAGGGGCTCTTGGTGATGAGCCAGCTGATCCGACAGGCT 4575  
 3994 TGCAATCCAGCTGCGAGAAACCTTTCTCTACTGTACACAGGGTGGAGTCTGTGTTGT 4053  
 4576 TGTAAACCCACCGTGTAGTCAACCCACTCGTACTGTAGCGAGACAAAACATGCCATGT 4635  
 4054 GAGAAGGGCTATACAGAGATATGAATCAATGGTTTCTCTGATTAATGCA 4105  
 4636 GAAGAAGGTTACTACTGAATCATGTTCTTAACAGCACCTTTGACGAATGCACACTTATC 4695  
 4106 -----TGAAGTACACAGGCTCAGAGGATAAAAGCTGATGTCAAAACCTTTCTG-G 4157  
 4696 CCGTGTGTGTTATCCACCACATGAGGAGCAAAAGAGGAGATGTGAACCAAGTCGGCT 4755  
 4158 GAAACACAGCTGTGAATCAAAATACATGATATTTTAAAGG--TGGTCTTTCAA 4215  
 4756 GTACATCCCAACCAACCCCTCCAGTAAACCCAGCAGGAGCGGGAAGACCTGTGTTCTACAG 4815  
 4216 CCACITGTATCCAGATGGCGAGTAAATAATTTGGGTTTATGGGTTTTCAGTGGCGCTTTT 4275  
 4816 CCATTGGCCAGATGGGAGACTAAGACCTGGTTTACGGTGTAGCAGCTGGGCAATT 4875  
 4276 CTCATCATGATTTTCTATATTTACTTCTACCTGTTTTCAGAGGCCA---AAACCA 4332  
 4876 GTGTACTCATCTTTTATTGTCTCCATGATTTATCTAGCTTGCAAAAGCCAAAGAACCC 4935  
 4333 CATCAAGCACACCTCCCAACAGAGCCTCTGACCTTAGCTACGATGAGGAGCTTAGAC 4392  
 4936 CAAGAGGCCAAACACCGACTGAACCTTTTAACTTAGCTATGATGGAGATGCCGAC 4995  
 4393 ATGTAA 4398  
 4996 ATGTAA 5001

## RESULT 6

CC72162

D ACC72162 standard; DNA; 6373 BP.

X C C C

X C C C

X T T T

X 07-JUL-2003 (first entry)

X Human NOV47a coding sequence.

X Human; NOV; antidiabetic; anorectic; antibacterial; virucide;

X immunomodulator; cytostatic; nootropic; neuroprotective;

X antiparkinsonian; antilipemic; gene therapy; metabolic disorder;

X diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;

X neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;

X immune disorder; haematopoietic disorder; gene; ds.

X Homo sapiens.

X S S S

X WO2003029423-A2.

X N X N

X D X D

X 10-APR-2003.

X X X X

02-OCT-2002; 2002WO-US031358.  
 02-OCT-2001; 2001US-0326483P.  
 05-OCT-2001; 2001US-0327342P.  
 09-OCT-2001; 2001US-0327917P.  
 09-OCT-2001; 2001US-0328029P.  
 09-OCT-2001; 2001US-0328044P.  
 09-OCT-2001; 2001US-0328056P.  
 12-OCT-2001; 2001US-0328849P.  
 15-OCT-2001; 2001US-0329414P.  
 17-OCT-2001; 2001US-0330142P.  
 22-OCT-2001; 2001US-0341058P.  
 24-OCT-2001; 2001US-0339286P.  
 24-OCT-2001; 2001US-0343629P.  
 29-OCT-2001; 2001US-0349575P.  
 01-NOV-2001; 2001US-0346357P.  
 12-APR-2002; 2002US-0371972P.  
 12-APR-2002; 2002US-0371980P.  
 17-APR-2002; 2002US-0373261P.  
 19-APR-2002; 2002US-0373805P.  
 23-APR-2002; 2002US-0374738P.  
 16-MAY-2002; 2002US-0381101P.  
 17-MAY-2002; 2002US-0381635P.  
 29-MAY-2002; 2002US-0383830P.  
 01-OCT-2002; 2002US-00262839.

(CURA-) CURAGEN CORP.

Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
 Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
 Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
 Rothenberg WE, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;  
 Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;

WPI; 2003-381625/36.

P-PSDB; ABR58450.

NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
 treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
 dyslipidemia, and in chromosome mapping, tissue typing or  
 pharmacogenomics.

Claim 20; Page 253-254; 487pp; English.

The present invention relates to novel human NOV proteins and their  
 coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
 proteins are useful in manufacturing a medicament for treating a syndrome  
 associated with a human disease. The NOV proteins and coding sequences  
 may be used to diagnose, treat or prevent metabolic disorders such as  
 diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
 disorders such as Alzheimer's disease or Parkinson's disease, immune  
 disorders, haematopoietic disorders and various dyslipidaemias

Sequence 6373 BP; 1836 A; 1385 G; 1603 G; 1548 T; 0 U; 1 Other;

Query Match

Best Local Similarity 28.0%; Score 1232.8; DB 7; Length 6373;

Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;

QY

10 TGCAATTCAGAGCTGACCGAAGTCTGGTTGCAATGAATGAATGCGACACTTTGCCCTT 69

Db

505 TGCATCCAGAAAGACAAAGACATCTCTGGGAGGATATCATCTGTGAGTACTTGAGCCC 564

QY

70 CAGCCTCTCTACAGAACAGGCTTGCCTCATTTCTTCCCGGGATTGTGTAGTATCTGAG 129

Db

565 AAGCCTCTCTCGAGCAGGCTTGCCTCATTTCTTCCAGCAGATTTGCATCGTGTCTGA 624

QY

130 TTCTTACCATGCTCCAACTGTAGCAGGATGTGGAGAGAAATTCAGCATAGAACTCGC 189

Db

625 TTTTGTGCTGTGTCGGAATGTCTCAAGACCTTCGGCAGCGGCTCCAGCACCGACGGT 684

QY

190 GCGGTCTACAGCTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGAGTCAAGA 249

685 CATGTGTGGGCCCCCGCGAGTTCTGGAGGCTCTGGCTGTCTCAAACTGTACGAGATTCCAG 744  
250 GCCTGTGATGCTCCCATTTCTCTCTCTTTGGGGAAGAGGAATATACATTTAGCCCTTAAG 309  
745 GTGTG-----CCAATCCAGTCCATGCGAGCGGAGGCTCAGGTACAGCCTGCA 795  
310 GTTGGACCATGGAGTAATGACAGACTGCTCATCTTAAGAAATTAATCAAGCGGAAGA 369  
796 GTGGGGCCCTGGAGCACTGCTCAATGCCCACTCCCGACANGTAAGACAAGCAAGAGA 855  
370 ACTGTTCTGGATTTTAA-----CTGTGATTCAAATGAG 402  
856 CGCGGAAGAATAAAGAACCGGAAGAGCCGAGCAAGAGGATTAAGGATCAGAGCC 915  
403 CGAGTCACTTTAAACATCAAGTTACA-----AAGCATCATCATCTCGAAGTCT 453  
916 CGGAGGCTTATTAGAAAAAGAGAAACAGAAACAGGCAAGACAGACAAGAGAACAAATAT 975  
454 TGGGCAATAGAGATAGGTATTCAAAACCGGCGAGTTTCGTGTACAGAAAGTGTGACAA 513  
976 TGGACATCCAGATTGGATATCAGCCAGAGAGTTATGTGCATTAACAAGCGGGGAA 1035  
514 AATGCTATGTTAAGCCTTTGCTTTCAAGATTCCTTCCCATTTGACTGTCTCCTGATC 573  
1036 GCTGCTGATTTAAGCTTTTGGCAAGAGAGAGCTTCCAAATGACCTTCCAGTCTCTGTG 1095  
574 ATGCCCAAAGACTGTGAACCTCCAGTGTCTCTCTGGAGCCCTGCTCCAAAGACATGC 633  
1096 ATCAACAAAGATGCCAGTTTCGAGTGTGTAGATGGAGCCCTGCTCAAAAACATGC 1155  
634 CGTTACAGGAGTCTCTTGCCAGGATTTAGAGCAGGAGCCGGAACGTGAGACATGGCT 693  
1156 CATGACATGTGTCCCTCGAGGCACTCGTGTAGGACAGCAACCATCAGGAGTTTCCC 1215  
694 ATTGGAGTGTGAAGAGTGTCTCTGACTTTCTGAGAAAGAGCCCTGCTATGTTAGGA 753  
1216 ATTGGAGTGTGAAGAGTGTCTCTGAGATTTTGAAGAAAAAGACCTGTTGTCTCAAGGA 1275  
754 GAATCTCTGCAGCAATCTCCAGGTATCTCTGGAGAACTTCTGAATGGAAAGAAATGCCAA 813  
1276 GATGGAGTTGTCCCTGTGCCAGTATGGCTGGAGAACTACAGAGTGGAGTGCCT 1335  
814 GTCTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGAGGGAACCGTGTGTGGC 873  
1336 GTGACCCCTTTGTCTAGTCAAGCAGGACAAGAGCGGCAACAGACGCGCCCTCTGTGGA 1395  
874 GGTGGGATCCAGACCCGGAGGTGTACTGTGCCAG-----AGCCTACCAAGCA 921  
1396 GGGGGCATCCAGACCCGAGAGGTGTACTGTGCGAGGCCAACGAACCTCTCTCACAA 1455  
922 GCTCCGCACTGAGGGCAAGGAAGTCTTAGACCTGTGGAAGCTGTGAAAAGGCAATTTATGTGGGA 981  
1456 TTAAAGTACCACAAGAACAAAGAGCCTCAAAGCCAAATGGACTTAAATTTATGCACTGGA 1515  
982 CCCGCCCTGTGCTCTCAGTCTGCAATATCCCTTGTCTACGAGCTGCATAGTATCT 1041  
1516 CCTATCCCTAATACACAGCTGTGCCATTTCTTGTCCAACTGAATGTGAAGTTCA 1575  
1042 TCCTGGTCAAGCTCGGCTGTGCATCCATCAAACTGTCAAACTCAGGGGAAAAA 1101  
1576 CTTTGGTCAAGCTCGGCTGTGACTTATGAAACTGTAACTGATCAGCAAGGAAAAA 1635  
1102 GGATTTAGNACGAGCGCG-----CCATGCTCTCATGNACTTACAGGCTGCA 1152  
1636 GCTTTCAAACTGAGGAAGCGGCGCATTTACCAATGAGCCCACTGGAGGCTCTGGGTAAAC 1695  
1153 GGGCAATTTGCCCTCATTTTGGTGGAGTCTGTTCTTGTGAGGATCCAAATGTCTACCGATGG 1212  
1696 GGAACATGCCCTCACTTACTTGAAGCACTTCCCTGTGGAAGCGCTGCTGTATGACTGG 1755  
1213 ---CTGGCATCAGAGGATCTGTTTCCCTGTATCATGGAA---AATGTGCCCTGGGACAT 1266  
1756 AAACGGGTGAGCTGGGAGACTCGGAGCCAGATAACGGAAAGGAGTGTGTCCAGGCAGC 1815

1267 CGTATTTCTGAAGGCGCTGTGCCAGAAATGACCGCGGAGAAGATGTATCAGGAGTCTTTGC 1326  
1816 CAAGTTCAAGAGGTTGTGTGATCAACAGTGTGAGAGAAGTTGACAGACACTGTGC 1875  
1327 CCAGTTCCTCCCTCTCTCTGAGAGGAAGTCTTGTGAAATTCCTTCCGGAATGGACTGTG 1386  
1876 AGAGATGCCATCTTCCCATCCCTGTGGCTGTGTATGCCCATGCCGGAAGACTGTGTG 1935  
1387 CTGAGCAGTGGAGGAGTGTGTATCTCTGTTCCTCCAGTCTCTTCAAATAAAAACTCAGAT 1446  
1936 CTGACACATGTTCTACGTGTCTCTCTCTGTCACACACCTGCTCAGGGAACACAGAA 1995  
1447 GGGAAACAGACAGGTCAAGAACTATCTCTGSC-----ACTGGCTGGGGAAGTGGAAAGCCA 1503  
1996 GGGAAACAGATACGAGCAGCATCCATTTCTGGCTATCGGGTGAAGAGTGGAAATTCGC 2055  
1504 TGTCCCTCTAGTCAAGGCTCTCCAAAGACATCGTTTGTGTATGACCATTCCTGTATGCAG 1563  
2056 TGTCCAAATGCAAGTCTTTCAGAGAGTACGAAGTGTATGAGCATCTTGTGCACAGTG 2115  
1564 CTTCACTGGGAGACATCCCTTGGGGCCCTTGTCTGAGAGACATTTGGTAATTCGCCCTT 1623  
2116 TACCACCTGCAAACTGTGTCCTTGGGGCCAGTGCATTTGAGGACACTCAGTATCTGCTTC 2175  
1624 AATGCAACCATTTGGCTGGAATGGAGAGCCAGCTGTGTGTAGGCATTTGAGACTCGAGA 1683  
2176 AACACACTACGACTTGGATGGGAGGCCCTCTGCTCTGTGCGCATGCGACAAGAAAA 2235  
1684 GTCTTCTGTCTCAAGAGTCACTGGGACAAGTAATGACCAAAAGATGTCCAGATTTCTACT 1743  
2236 GTCACTCTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAAAATGCTCTGAAAGCCTT 2295  
1744 CGACTGTAATGTGGCCCTCTGTTTCTCCATGCAAAAAGACTCTTATGTGACTGCT 1803  
2296 CGACTGTAATGTGAAGCCTTGTCTCTTGTGAAGAGGACTGTATTTGACCCCA 2355  
1804 TTCACTGAGTGGACACCTCTGCCAAGGATGTGCAAGCAGGAAATGCCACAGT---AAAA 1860  
2356 TATAGTACTGGACATCATGCCCCCTCTCTGTTGTAAGAGGGGACTCCAGTATCAGGAAG 2415  
1861 CAGTCTCATACAGATCATCATCAAGAGACGCAATGAGAGCCAGGATGCGCAGAT 1920  
2416 CAGTCTAGGCACTGGGTCTCATCACTGCTGCCAACCGGGGCCGAGACTGCACAGAT 2475  
1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTTGTCTCTGTATATCGGTGGAAG 1980  
2476 CCCCTCTATGAGAGAGGCTGTGAGGCACCTCAAGCTGCCAAAGCTACAGTGGAG 2535  
1981 CCACAGAAATGAGGCCCTTGTGATCTTATGTCAGAGTCTGTCTGGCAGGGAATTAACGGGC 2040  
2536 ACTCAAAATGGCGCAGATGCCAATTTAGTCCCTTGGAGCGGTGCAACAGACAGCCCTGGA 2595  
2041 AGCAGTGAAGCCTGTGGAAGGGGTACAAACAAGAGTGTCTCATGCTCTCTGATGAC 2100  
2596 GCACAGGAGGCTGTGGCCTTGGCGCAGCAGCAGAGCCATTACTTTGCGCAACAGAT 2655  
2101 AACCGGTGAGCAAAATGATGGAATGCTCAAGCAGACAAACGGCATGCTCTCTCTTGTG 2160  
2656 GGAGGACAGGCTGGAATTCATGAGTGTGCTACAGTATGAGGCCCTGTGCCAGCCCTTACC 2715  
2161 CAAGAATCCACAGTCCCATGTGCGAGAGAGTGCACCTTCACTGCTGTCTCAAGTTCACG 2220  
2716 CAGGCTGCCAGATCCCTGTCAGAGTACTGTCAATTTGACAGCTGCTCCAGTTTTCT 2775  
2221 CCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGTGGGGGAGCAGCTCACAGGGAAGC 2280  
2776 TCATGCAATGAGACTGTGTGCGAGTTAGGACAGAAAGCGCACTCTTTGTTGGAAGAAGT 2835  
2281 AGAAGAAGGAGAAATGCCAGGATTTGACCTTTACCTTTAGTGGAGACAGACTATGT 2340  
2836 AAAAGAAGGAAAAATGTAAAAATTTCCATTTGTATCCCTGATTTGAGACTCAGTATGT 2895

2341 CCTTGTGATGAATTTATATCCCAACCTTATGGAACCTGTCAGATTGCATTTCTTCAGAA 2400  
2896 CCTTGTGACAAATATATGACAAACCTGTGGGAACTGTCAGACTGATTTTACCAGAG 2955  
2401 GGCAGAGGAGGACCTCACCGAGGCTGCGGTACAGCAGACAGCAAGAAATGTGGAGAA 2460  
2956 GGAAGAGTGGAGTGTGCTGGGAATGAAAGTACAGGAGACATCAAGGAATTCGGGACAA 3015  
2461 GGCCTCGCTTTTCGAGCAGTAGCTTCTGTGATAAAATGGAAGACCTGTGACCCCTCC 2520  
3016 GGATATCGTTACCAAGCAATGGCATGCTACGATCAAAATGGCAGGCTGTGGAACATCT 3075  
2521 TTCTGACGACGCTCTGGTTACATTAAGAAAAATGTGTCAITTCCTGCCCATTTGATTCG 2580  
3076 AGATGTAACAGGCAATGGTTACATTTGAGAGGCTGTCATCATCCCTGCCCTCAGACTGC 3135  
2581 AAGTTAAGCGATGTGTCTAGTTGGGGCTCTTCAGTTTCATCTTGTGGAAATGAGGTGAGA 2640  
3136 AAGCTCAGTGTGTCTCAACTGGTGGCTGCGCTGCGCAAGTCTCTGTGGAGTGGTGTGAAG 3195  
2641 ATTTCGATCCAAATGGGTAAAGAAAAACCTTCAATGGAGGACGACCAATGTCCTCCAACTG 2700  
3196 GTTCGTCTTAAATGGCTGGTGAAGAAACCATATAATGGAGGAAGGCTTGTCCCCCAACTG 3255  
2701 GATC-----TCAGAAATCAGTACATGAGGAGTCCCATGTTACAGTGTGCAATCAG 2754  
3256 GACCATGTCAACAGGCAAGGTGTATGAGGTGTTGCCATGCCAGTGTGCAACCCAG 3315  
2755 TATTCCTGGGTTGTAGAACACTGGTCTTCATGCAAAATCAACAAATGAGTGTGAGTCCCTG 2814  
3316 TACCTATGGGTACAGAGCCCTGGAGCATCTCAAGGTGACCTTGTGAATATGGGGAG 3375  
2815 CGCTGTGGAGGAGGACAAATCTAGGAAATCAGATGTGT---GAATCTCGGATGGT 2871  
3376 AACTGTGGAGAGGGCGTGCAAACCCGAAAGTGTGATCTGATGATGATGATGATGATGATG 3435  
2872 GAAGGTGGAGCAGTGTAGTACCACTGTGCAACAGGATGAAATCCCCAGAAACCCAG 2931  
3436 CCTTCTGAACATGTAGAGGATACCTCTGTGACCCAGAGATGCCCCCTGGCTCTAGA 3495  
2932 TCTGTCTTCTATGTGTCCCAATGATGTGTATGTGTGAGTGGGACTTTGGAGCAA 2991  
3496 GTGTGCAAAATACCATGCTGAGGACTGTGTATCTGATGATGATGATGATGATGATGATG 3555  
2992 TGCCACAGTGTGATGCTCCACACAAAT---GCAGAGAAAGTGTGATGATGATGATGATGATG 3048  
3556 TGTTTTCCTTGCATCAAGCAGTTTCCGCAAGGTGAGTGTGATGATGATGATGATGATGATG 3615  
3049 CCATCACTGAACCTCAAGGACTTTGTGTGAAGACTCACAGGTGACGCTTGCCTCTGAT 3108  
3616 CCAGCTGATGAAGGAAGATCTTGCCTTAATGCTGTTGAGAAAGACCTGTAACTGTAAC 3675  
3109 GAAATTTGCTTCCAGTTCAGTACATCTAACAGAGTGGAGCAGATGCCAGCTGATGAA 3168  
3676 AAAAACTGTACCACTATGATTTATATGTAAGTGAAGTGTGATGATGATGATGATGATGATG 3735  
3169 AACGCACTGTGCTCAAGGCGTCAAGGACCCGCTGCTAAGCTGTGTGTGAGTGTGATGCTG 3228  
3736 AAGGCACTTTGTGGAAATGGAATMAAACCAAGATGTTGGATTGTTGCAAGTGTGATGCTG 3795  
3229 AAGCCAGTCAAGATGGAACCAATGTGAGCAGCAATATTGGAGAACCCCAAGAAATGAGC 3286  
3796 AAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAACTGGCAGATGAAC 3855  
3289 ATTCCCTGCTTGGTGAATGCGTGTGCACTGTGAGTCTCAGGCTGAGCGCTTGGACA 3348  
3856 AGCTCTGATGATGGAATGCTGTGAACTGTGAGCTTCTGATGATGATGATGATGATGATG 3915  
3349 GAGTGTTCACAGCTGTGGCCATGAGGTGCAATGAGCCGACTCGATTTATCATATTATG 3408  
3916 GAATGTCTTCAAAATGTTGGCTCTCACAGGAAATGATCCGAAGACGCAAGTGTGACCCAG 3975  
3409 CCAACCCAGGAGAGGAGCGGCTATGCCCCACAGAGCTTACCAGGAGAAACCTTGCCCA 3458

Db CCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGAGCAGACCTCCAAACCTGCCCA 4035  
QY GTGACCCCTCTGTACAGCTGGTCCCTTGGCAACTGTGTCTGCATGTAAATTTGAGAGGTGGA 3528  
Db GTGAAGCTTTGTTATCGGTGGCAATATGSCCAGTGTCTCCATSCCAAGTSCAGGAGGCC 4095  
QY GACTGTGGGGAAGAGTTCAGATCGCAGCCTTTCTCTGCAATGTCACAGTGTGTTCAATA 3588  
Db CAGTGTGGAGAGGGACCAGAAACAAGAACATTTCTGTGTAGTAAAGTGTGAGTGTGAGT 4155  
QY TCTCATGCACTGGAACGTGTGAGGATGCACTGTGTGGAGAAATGCCCTT----- 3638  
Db GATGATTTCAAGCAAGTGTGTGATGAGGAATTTCTGTCTGACATTTGAACCTCATATAGAT 4215  
QY -----TCAGAGACAGCATCTTGAAGCAGCTGTGTCTGTGCTTGGCCAGGAGCTGCCAT 3693  
Db GGTATATAAATAATGTGTTCTGGAGGAATCTCTGACGAGCTTGGCCAGGTGACTGTAT 4275  
QY TTAACAGAAATGTCAGAGTGGAGCAGATGTGAATTAACCTGCAATTTGATGGAAGAAGCTTT 3753  
Db TTGAAGGACTGGTCTTCTCGAGCCTGTGTGTCAGTGCCTGTGTGATGATGATGATGATGAT 4335  
QY GAGACTGTGGCCCGCCAGTCTAGATCAAGCACTTTTATATTCAGTCTTTTGTGAGAACAA 3813  
Db GGCTTTGGTGGAAATACAGGTCTGATCCAGACCGGTGATTTATACAGAACTTAGAGATCAG 4395  
3814 GACAGCTGCCCCCAACAGAGTCTTAGAAACACGCGCTTGTACAGAGCGCAATGTTATCAC 3873  
Db CATCTGTGCCCAGAGCAGATGTTAGAAACAANAATCATGTTATGATGACAGTCTATGAA 4455  
QY TACATGTGAAGAGCAAGTCTTTTGGAAACATAACAGAACTGTATGTTGGTCCAGCGTTCA 3933  
Db TATAATTGGATGGCCAGTGTCTTGGAAAGGGCTCTTCCGAAACAGTGTGTGTCAAGGTCA 4515  
QY GATGGCTTAATGTCAAGGAGCTGTCTCCCTCAGCCCGCTCTGCTGCTGCTGCTGCTGCTG 3993  
Db GATGTTATAATGTACAGGGGCTGTGTTGTTGATGATGATGATGATGATGATGATGATGATG 4575  
3994 TGCATTCAGCCTTCAGAAACCTTTCTCTACTGTACAGAGGTGGAGTCTGTGTTGT 4053  
Db TGTAAACCCACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAACATGCCATTGT 4635  
QY GAGAAGGCTTATCAGAGATAATGAATCAAAATGTTTCTCTGATTTACTGCA----- 4105  
Db GAAGAGGGTACACTGAAGTCTGTTCTTAACAGCACCCCTTGAGCAATGACACTTATC 4695  
4106 -----TGAAGTACCCAGGCTCAGAGGATAAAAAAGCTGATGAAAAACCTTTCTG-G 4157  
Db CCGTGTGTGTTATCCCACTATGGAGGACAAAGAGGAGATGTGAACCCAGTCCGGCT 4755  
QY GAAAAACAGACCTGTGMAATTCAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
Db GTACATCCAAACCAACCCCTCCAGTAAACCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4815  
4216 CCATTTGATCCAGATGGCCGAGTAAATAATTTGGGTTTATGGCGTTTTCAGGTGGCGCTTTT 4275  
Db CCATTTGGGCGAGTGGAGACTTAAGACTGCGGTTTACCGTGTAGAGCTGGGGCATTT 4875  
QY CTGATCATGATTTTCTTAAATTTTATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4332  
Db GTGTTACTCATCTTTATTTGTTCTCCATGATTTATCTAGCTTGTGAAAAAGGCAAAAGAACCC 4935  
4333 CATCAAGACACCTCCCAACAGAGGCTCTGACCTTACCTACCTAGGAGCTTAGAC 4392  
Db CAAAGAGGCAAAACCAACCGACTGAAACCTTTAACTTACCTTATGATGAGATGCCGAC 4995  
QY ATGTAA 4398  
Db ATGTAA 5001

RESULT 7



ACC721171  
ID ACC721171 standard; DNA; 6373 BP.  
XX  
XX ACC721171;  
XX  
XX ACC721171;  
DT 07-JUL-2003 (first entry)  
XX  
XX Human NOV47J coding sequence.  
DE  
XX  
XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytostatic; nootropic; neuroprotective;  
KW antiparkinsonian; antilipemic; gene therapy; metabolic disorder;  
KW diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; gene; ds.  
XX  
OS Homo sapiens.  
XX  
XX W02003029423-A2.  
PN  
XX  
XX 10-APR-2003.  
PD  
XX  
XX 02-OCT-2002; 2002WO-US031358.  
PF  
XX  
XX 02-OCT-2001; 2001US-0326483P.  
PR  
XX 05-OCT-2001; 2001US-0327342P.  
PR  
XX 09-OCT-2001; 2001US-0327917P.  
PR  
XX 09-OCT-2001; 2001US-0328029P.  
PR  
XX 09-OCT-2001; 2001US-0328044P.  
PR  
XX 09-OCT-2001; 2001US-0328056P.  
PR  
XX 12-OCT-2001; 2001US-0328849P.  
PR  
XX 15-OCT-2001; 2001US-0329414P.  
PR  
XX 17-OCT-2001; 2001US-0330142P.  
PR  
XX 24-OCT-2001; 2001US-0341058P.  
PR  
XX 24-OCT-2001; 2001US-0339268P.  
PR  
XX 24-OCT-2001; 2001US-0343628P.  
PR  
XX 29-OCT-2001; 2001US-0349575P.  
PR  
XX 01-NOV-2001; 2001US-0346357P.  
PR  
XX 12-APR-2002; 2002US-0371972P.  
PR  
XX 12-APR-2002; 2002US-0371980P.  
PR  
XX 17-APR-2002; 2002US-0373261P.  
PR  
XX 19-APR-2002; 2002US-0373805P.  
PR  
XX 23-APR-2002; 2002US-0374738P.  
PR  
XX 16-MAY-2002; 2002US-0381101P.  
PR  
XX 17-MAY-2002; 2002US-0381635P.  
PR  
XX 29-MAY-2002; 2002US-0383830P.  
PR  
XX 01-OCT-2002; 2002US-00262839.  
XX  
XX (CURA-) CURAGEN CORP.  
PA  
XX  
XX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
PI Rothenberg ME, Shimkots RA, Smithson G, Spytek KA, Taupier RJ;  
PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
XX  
XX WPI; 2003-381625/36.  
DR  
XX P-PSDB; ABR38459.  
XX  
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
PT dyslipidemia, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX  
XX Claim 20; Page 263-264; 487pp; English.  
PS  
XX  
XX The present invention relates to novel human NOV proteins and their  
CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
CC proteins are useful in manufacturing a medicament for treating a syndrome  
CC associated with a human disease. The NOV proteins and coding sequences  
CC may be used to diagnose, treat or prevent metabolic disorders such as  
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
CC disorders such as Alzheimer's disease or Parkinson's disease, immune



QY	3109	GAAAAATGCTTCCAGTTCAGTACAATCTTAACAGAGTGGAGCACATGCCAGCTGAGTGA	3168
Db	3676	AAAAACCTGCTACCACCTATGATTAATAATGTAACAGACTGGAGTACATGTCACTGAGTGA	3735
QY	3169	AACGCACCCCTGTGTCAAGCGCTCAGGACCGGCTGCTAAGACGTGTGTGTCAGTGATGC	3228
Db	3736	AAGCAGTTTGTGGAAATGGAATAAANAACAAGATGTTGGATTGTGTTGCAAGTGAATGC	3795
QY	3229	AAGCCAGTCAGCATGGACCAATGTGACGACATAATTTGGAGAAGCCCCAGAGAAATGAGC	3288
Db	3796	AAGTCAGTTCCACCTGAAATATTGTGAAGCCCTTGGCTTGGAGAAGAACCTGGCAGATGAAC	3855
QY	3289	ATTCCTGCTTGTGGAAATCGGTGCTCAACTGTCACTGTCACTGAGGTGGACGGCTTCGACA	3348
Db	3856	ACGTCCCTGCAATGTTGGAAATGCCCTGTGAATGTCAAGCTTTTCAATTGGTCTCCTTGGTCA	3915
QY	3349	GAGTGTTCACAGACCTGTGSCCATGGAGGTGCAATGAGCCGGACTCGATTTATCATTTATG	3408
Db	3916	GAATGTTCTCAAAACATGTGGCCCTCACAGGAAAAAATGATCCGAAGACGAACAGTGAACCCAG	3975
QY	3409	CCAACCAAGAGAGAGACGGCCATGCCACAGAGGCTTACCAGGAGAGAAAACCTGCCCA	3468
Db	3976	CCCTTCAAGTGTATGAGAGACCATGCCCTTCCCTGATGGAACAGCTCCAAACCTGCCCA	4035
QY	3469	GTGACCCCTGCTACAGCTGGTCCCTTGGCAACTGGTCTGCATCTATAATTTGGAGGTTGA	3528
Db	4036	GTGAAGCCTTGTATTCCGTTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCC	4095
QY	3529	GACTGTGGGGAAGGAGTTCAGATCCGACGCTTTCCTGCACTGTCCACAGTGGTTCAATA	3588
Db	4096	CAGTGTGGAGAGGGAACAGAAACAAGGAAACATTTCTTGTGTAGTGAATGGGTGAGCT	4155
QY	3589	TCTCATCAGCTGGACGTGTGAGGATGTCACCTGTGTGAGAAAAATGCCCTT-----	3638
Db	4156	GATGATTTACGAAAGTGGTGGATGAGGAATCTGTGCTCACAATTGAATCACTATTATAGAT	4215
QY	3639	-----TCAGGACAGCATCCTGAAGCAGCTGTGTCTGTGCTTGGCCAGAGACTGCCAT	3693
Db	4216	GGTAATAAAATATGTTTCTGGAGGAATCTGCAGCCAGCCTTGGCCAGGTGACTGTTAT	4275
QY	3694	TTAACAGAAATGGTCAGAGTGGAGCACATGTGAATTAACCTGCCATTTGATGAAGAAAGCTTT	3753
Db	4276	TTGAAGGACTGGTCTTCCTGAGCCCTGTGTCACTGACCTGTGTGAATGTGTGAGGATCTA	4335
QY	3754	GAGACTGTGGCCGCCAGTCTAGATCAAGACATTTTATAATTCAGTCTTTTGAGAACCAA	3813
Db	4336	GGCTTTGGTGGAAATACAGGTTCAGATCCAGACCCGGTGATTTACAAGAACTAGAGAAATCAG	4395
QY	3814	GACAGCTGCCCCCAACAGGTTCTAGAAACACGCCCTTGTACAGGAGCAAAATGTTATCAC	3873
Db	4396	CATCTGTGCCAGAGCAGATCTTAGAACAATAATCATGTTATGATGACAGTGTCTATGAA	4455
QY	3874	TACACATGGAAAGCAAGTCTTTTGAAACAAATAACGAACTGTATGGTCCGACGGTTCA	3933
Db	4456	TATAAATGGATGGCCAGTCTTGGAAAGGGCTCTTCCGAAACAGTGTGGTGTCAAAGGTCA	4515
QY	3934	GATGGCTTAATGTACAGGAGGCTGTGCCCTCAGGCCCGTCTGTGCTGCCATTCGCGCAG	3993
Db	4516	GATGGTATAATGTAAACAGGGGCTGTTGGTGTATGAGCCAGCCTGATGCCGACAGGTCT	4575
QY	3994	TGCATTCAGCCTTCAGAAAACCTTTTCTCCTACTGTHACAAGGGTGGAGTCTGTGGTTGT	4053
Db	4576	TGTAAACCAACCGTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAAACATGCCATTGT	4635
QY	4054	GAGAAGGGCTATACAGAGATAATGAAATCAAAATGGTTTCTCTGGATTACTGCA-----	4105
Db	4636	GAAGAAAGGTTACATGAAGTCACTGTCTTCTTAACAGCACCTTGGACCAATGCACATTATC	4695
QY	4106	-----TGAAAGTACAGGCTACAGGATATAAAAGACTGATGTGAAAAACCTTTCTG-G	4157
Db	4696	CCCGTGTGTGTTATTAACCAACATGGAGGACAAAGAGAGAGATGTGAAAAACAGTCCGGCT	4755

Qy	4158	GA	AAACAGACGCTGTGAATTC	AAAAAATACATGATATTTT	TAAGGA--TGGTCTCTTCAA	4211
Db	4756	GT	ATATCCAACCAACCCCTCC	AGTAACCCAGCAGGAGGAC	CTGGTTTCTTACAG	4815
Qy	4216	CC	ACTTGTATCCAGATGGCGAG	TAAATAATTCGGGTTTAT	TGGCGTTTCAGGTGGCGCTTTT	4275
Db	4816	CA	TTTGGCCAGATGGGAGACT	TAAGACCTGGGTTACGGT	TGTACGACTGGGCAATTT	4875
Qy	4276	CT	CATCATGATTTTCTTAATAT	TATTTACTTCTCCTGATTT	ATCTATCTAGCTTGA	4332
Db	4876	GT	GTACTCATCTTTTATTGTC	TCTCCATGATTTATCTAT	CTAGCTTGA	4935
Qy	4333	CAT	CAAGCAGACACCTCCCAAC	CAGAGCCTCTGACCTTAG	CCCTACGATCGAGACTTAGAC	4392
Db	4936	CA	AGGAGCGAATAACACCGAC	TGAAACCTTTAA	CCCTTAGCCTATGATGAGATGCCGAC	4995
Qy	4393	AT	GTA	A	4398	
Db	4996	AT	GTA	A	5001	
RESULT 8						
ACC72168						
ID ID ACC72168 standard; DNA; 6373 BP.						
XX	AC	ACC72168;				
XX	AC	XX				
XX	AC	XX				
DT	07-JUL-2003	(first entry)				
XX	XX	Human NOV47g coding sequence.				
DE	XX	Human; NOV; antidiabetic; anorectic; antibacterial; virucide;				
XX	XX	immunomodulator; cytostatic; nootropic; neuroprotective;				
KW	KW	antiparkinsonian; anlipaemic; gene therapy; metabolic disorder;				
KW	KW	diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;				
KW	KW	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;				
XX	XX	immune disorder; haematopoietic disorder; gene; ds.				
XX	OS	Homo sapiens.				
XX	XX	WO2003029423-A2.				
XX	XX	10-APR-2003.				
PD	XX	02-OCT-2002; 2002WO-US031358.				
PF	XX	02-OCT-2001; 2001US-0326483P.				
XX	XX	05-OCT-2001; 2001US-0327342P.				
PR	PR	09-OCT-2001; 2001US-0327917P.				
PR	PR	09-OCT-2001; 2001US-0328029P.				
PR	PR	09-OCT-2001; 2001US-0328044P.				
PR	PR	09-OCT-2001; 2001US-0328056P.				
PR	PR	12-OCT-2001; 2001US-0328849P.				
PR	PR	15-OCT-2001; 2001US-0329414P.				
PR	PR	17-OCT-2001; 2001US-0330142P.				
PR	PR	22-OCT-2001; 2001US-0341058P.				
PR	PR	24-OCT-2001; 2001US-0339266P.				
PR	PR	29-OCT-2001; 2001US-0343629P.				
PR	PR	01-NOV-2001; 2001US-0349575P.				
PR	PR	12-APR-2001; 2001US-0346357P.				
PR	PR	12-APR-2002; 2002US-0371972P.				
PR	PR	12-APR-2002; 2002US-0371980P.				
PR	PR	17-APR-2002; 2002US-0373261P.				
PR	PR	19-APR-2002; 2002US-0373805P.				
PR	PR	23-APR-2002; 2002US-0374738P.				
PR	PR	16-MAY-2002; 2002US-0381101P.				
PR	PR	17-MAY-2002; 2002US-0381635P.				
PR	PR	29-MAY-2002; 2002US-0383830P.				
PR	PR	01-OCT-2002; 2002US-00262839.				
XX	XX	(CURA-) CURAGEN CORP.				
XX	XX	Alisoebrook JP; Anderson DW; Boldog FL; Burgess CE; Catterton E;				

I Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
I Kekuda R, Leach MD, Li L, Miller CB, Patturajan M, Rieger DK;  
I Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;  
I Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
X WPI; 2003-381625/36.  
X P-PSDB; ABR58456.  
X NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
T treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
T dyslipidemia, and in chromosome mapping, tissue typing or  
T pharmacogenomics.  
X Claim 20; Page 260-261; 487pp; English.  
X The present invention relates to novel human NOV proteins and their  
C coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
C proteins are useful in manufacturing a medicament for treating a syndrome  
C associated with a human disease. The NOV proteins and coding sequences  
C may be used to diagnose, treat or prevent metabolic disorders such as  
C diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
C disorders such as Alzheimer's disease or Parkinson's disease, immune  
C disorders, haematopoietic disorders and various dyslipidaemias  
X Sequence 6373 BP; 1836 A; 1385 C; 1603 G; 1548 T; 0 U; 1 Other;  
Query Match 28.0%; Score 1232.8; DB 7; Length 6373;  
Best Local Similarity 57.7%; Pred. No. 0;  
Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;  
Y 10 TCATTCAGAGCTGAACGAACTGTGGTGTGCAATGAAATATGCAACATTTCGCCCTT 69  
b 505 TGCATCCAGAAAGACAAAGACATTCCTCGGAGGATATCACTGTGAGTACTTTGAGGCC 564  
Y 70 CAGCGCTCTACAGACAGCTGCTCATCTCTGCTCCCGGGATGTGTAGTATCTGAG 129  
b 565 AAGCCTCTCTGAGCAGCTTGGCTCTCACTTCTGCGAGCAAGATTGCATCGTCTGAA 624  
Y 130 TTCTTACCATGGTCCAACTGTAGCAAGGGATGTGGGAAGAAATTCGACATAGAACTCGC 189  
b 625 TTTTCTGCTGGTCCGAATGTCTCAAGACCTCGCGCAGCGGCTCCAGCCGCGGT 584  
Y 190 GCGGTCATAGTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGATGATCTCAAGA 249  
b 685 CATGGTGGCGCCCGCCAGCTTGGAGGCTCTGCGCTGTCCAAACCTGACGGAGTTCCAG 744  
Y 250 GCGTGTGATGCTCCCATTTCTCTGCTCTTGGGAAGAGGATATACATTTAGCCTTAAG 309  
b 745 GTGTG-----CCAATCCAGTCCATGCGAGCGCCGAGGCTCAGGTACGCTGCAT 795  
Y 310 GTTGACCATGGATGAATGCAAGCTGCTCATCTTAAAGAAATTAATCCAGCGGAAGA 369  
b 796 GTGGGGCCCTGGAGCACCTGTCTCAATGCCCCCACTCCCGACAAAGTAAAGCAAGAGGA 855  
Y 370 ACTGTCTCGAATTTAA-----CTCTGATTCGAATGAG 402  
b 856 CCGGGAAGATTAAGACCGGAAGAGCCGAGCAAGAGGATTAAGGATCCAGAGGCC 915  
Y 403 CGAGTCACTTTAAACATCAAGTTTACA-----AAGCACATCATCATTCGAAGTCT 453  
b 916 CCGGAGCTTATTAGAAAAAGAGAAACAGAAACAGGACAGACAGCAAGAGAAACAAATAT 975  
Y 454 TGGCAATAGAGATAGGTTATCAACCCGCGAGGTTCTGTGTACAAAGATGATGGACAA 513  
b 976 TGGGACATCCAGATTGGATATCAGACCAAGAGGTTATGTGCAATTAAAGAGCGGGAAA 1035  
Y 514 AATGCTATGTTAAGCCTTTGCTTCAAGATTCCTTCCCATGACTGTTTCACTCTCGCATC 573  
b 1036 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGAGCTTCCAAATGACTTCCAGTCTGTGTG 1095  
Y 574 ATGCCCAAGAGCTGGAACCTCCAGTGGTCTCTCTGGAGCCCTGCTCAGAGATGC 633  
b 1096 ATCAACAAAGAGTGCAGGTTTCCGAGTGGTGCAGATGGAGCCCTCTCTCAAAAACATGC 1155

QY 634 CGTTCAGGAGTCTCTTGGCAGGATTTAGAGCAGGAGCGGAACCTGAAGCACATGGCT 693  
Db 1156 CATGACATGTTGTTCCCTCGAGGCATCTGTGTAAAGACAGACACATCAGGAGTTTCC 1215  
QY 694 ATTGGAGTGTGAAGGAGTGTCTTGAACCTTTGAGAAAGAGCCCTGCAATTGTTGAAGA 753  
Db 1216 ATTGGCAGTGAAGGAGTGTCCAGAAATTGAGAAAAAGAACCTGTTTGTCTCAAGGA 1275  
QY 754 GAACTTCTGCAGCAATGTCCAGGATTTCTCTGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 1276 GATGGAGTGTCCCTGTGCCAGTATGCTGTGAGAACTACAGAGTGGACTGAGTCCCGT 1335  
QY 814 GTCTCTCTCTCTCGAGCAGGATCCCCACTGTGCATGTGACGGGACCCCTGTGTGGC 873  
Db 1336 GTGGACCTTTGCTCAGTCAGCAGGACAAGAGCGGGCGCAACAGACAGCGCCCTCTGTGA 1395  
QY 874 GGTGGGATCCAGACCCGGAGGTGTACTGTGCCAG-----AGCGTACCAGCA 921  
Db 1396 GGGGGCATCCAGACCCGAGAGGTGTACTCGGTGAGGCCCAAGAAACCTCTCTCACAA 1455  
QY 922 GCTGCCCACTGAGGGCCCAAGGAAGTCTCTAGACCTGTGAAAAAGGCATTTATGTGGGA 981  
Db 1456 TTAAGTACCCACAAGAAACAAAGAGCCTCAAAGCCAATGCACTTAAATATGCACTGA 1515  
QY 982 CCGCGCCCGTGGCCCTCTCAGCTCTGCAATATCCCTTGTCTTACGACTGCAATGATCT 1041  
Db 1516 CCTATCCCTTAATACACAGCTGTGCCACATTCCTTGTCCAACCTGAATGTGAAGTTTCA 1575  
QY 1042 TCTGTGTGAGCTGGGGCCTGTGCATCCATGAAACTGTGCATGAACCTCAGGGGAAAAA 1101  
Db 1576 CCTTGTGACCTGGGACCTTGTACTATGAAACTGTAAATGATCAGCAAGGAAAAA 1635  
QY 1102 GATTTAGAACAGGAGCG-----CMAATGCTCATGGAATCTACAGGGCTGCA 1152  
Db 1636 GGGTTCAAACTGAGGAAGCGGGCATTTACCAATAGCCCCACTGGAGGCTCTGGGGTAACC 1695  
QY 1153 GGGCATGGCCCTCATTTGGTGGAGTCTGTTCTGTGAGGATCCCAATGCTACCGATGG 1212  
Db 1696 GMAACTGGCCCTCACTTACTGTGAGCCATTCCTGTGAGAGCCTGCTGTATGACTGG 1755  
QY 1213 ---CTGGCATCAAGAGGATCTGTTTCCCTGTATCATGGA---AATGTGGCTGGGACAT 1266  
Db 1756 AAGCGGTGAGACTGGGAGACTCGGAGCCAGATAACGGAAGAGGTGTGTCAGGACG 1815  
QY 1267 CGTATTTCAAGGCGCTCTGCCAGATGACCGGAGAGATGTATCAGGAGTCTTTGC 1326  
Db 1816 CAAGTTCAAGAGGTTGTGTGCATCAACAGTGTAGAGAGAGTTGACAGACAGCTGTGC 1875  
QY 1327 CCAGTTCCCGCTCCTCTGAGAGGAGTCTTGTGAAATTTCCCTGCCGAATGGACTGTGTG 1386  
Db 1876 AGAGATGCCATCTTCCCATCCCTGTGGCCTGTGATGCCCATGCCGAAAGACTGTGTG 1935  
QY 1387 CTGAGCGAGTGAAGGAGTGTCTCTGTTCCAGTCTCTGTTTCAAAATAAAACCTCAGAT 1446  
Db 1936 CTCAGCACATGTTCTACGTTGCTCTCTGCTCACACCTGCTCAGGGAAAAACGACAGAA 1995  
QY 1447 GGGAAACAGACAGGCTCAAGAACTATCTTGGC---ACTGGCTGGGCGAAAGTGGAAAGCA 1503  
Db 1996 GGGAAACAGATCAGGACAGATCCATTCTGGCTATCGGGTGAAGAGGTGGAATTCG 2055  
QY 1504 TGTCCCGCTAGTCAGGCTCTCAAGAGCATCTTTGTGTATGACCATTCCTGTATGAG 1563  
Db 2056 TGTCCAAATAGCAGTGTCTTGAAGAGTACGAAGCTGTAAATGAGCATCTCTTGACAGTG 2115  
QY 1564 CTTTACTGGGACATCGCCTTGGGGCCTTGTCTGAGGACACATTTGTTACTCCCTT 1623  
Db 2116 TACCAGTGGCAACTGGTCCCTTGGGGCCAGTGCAATGAGACACCTCAGTATCTGCTTC 2175  
QY 1624 AATGCAACATTTGGTGAATGAGGAAGCCACGTTGGTGTAGGCAATTCAGACTCGGAGA 1683  
Db 2176 AACACAACTACGACTTGAATGGAGGAGGCTCTCTCTCTGCGGATGCAGACAAAGAAA 2235

QY 1584 GTCTTCTGTCTCAAGATCTAGTGGGCAAGTAATGACCAAAAGATGTCCAGATTCTACT 1743  
Db 2236 GTCACTCTGTGTGCGAGTCAATGTGGGCAAGTGGGACCCCAAAAATGTCTGAAAGCCTT 2295  
QY 1744 CGACCTGAAACTGTGCGCCCTGTTTCTCCCATGCAAAAAGACTGTATTGTGACTGCT 1803  
Db 2296 CGACCTGAACTGTAAAGCCTTGTCTGCTTCTTGTGAAGAGACTGTATTGTGACCCCA 2355  
QY 1804 TTCAGTGTAGTGGACACCCCTGCCAAGGATGTGCAAGCAGAGAAATGCCACAGT---AAAA 1860  
Db 2356 TATAGTGTACTGGACATCATGCCCTCTTCTGTGTAAAGAGGGGACTCCAGTATCAGGAG 2415  
QY 1861 CAGTCTCGATACAGAAATCATCCRAAGAGCAGCCAAATGAGGCCAGGAATGCCAGAT 1920  
Db 2416 CAGTCTAGGCATCGGGTCAATCATTCAGCTGCCAGCAACGGGGCCGAGACTGCACAGAT 2475  
QY 1921 ACCTTATATAGGAGAGAGAGTGTGAAGATGTTTCTTGTGTCTGTATATCGGTGGAAG 1980  
Db 2476 CCCCTCATGAAGAGAGGCGTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGGTGGAAG 2535  
QY 1981 CCACAGAAATGGAGCCCTTGCACTTCTAGTCCAGAGTCTGTCTGGCAGGGAATTAACGGGC 2040  
Db 2536 ACTCACAATGTGGCCAGATGCCAATTAGTCCCTTGGAGCGTGCACAAGACAGCCCTGGA 2595  
QY 2041 AGCAGTGAAGCCTGTGGAAGGGGTTACAAACAGAGCTCTCTCATGCATCTCTGTATGAC 2100  
Db 2596 GCACAGGAAGCCTGTGGCCCTGGGCGACAGGAGGCCATTACTTGTCCAGCAAGAT 2655  
QY 2101 AACCGGTCCAGAGAAATGATGGAATGCTCTCAAGCAGACAAACGGCATGCTCTCTTGTG 2160  
Db 2656 GGAGGACAGGCTGGAAATCCATGATGCTCTACAGTATGCAGGCCCTGTGGCAGCCCTTACC 2715  
QY 2161 CAAGAATGCACAGTCCCATGTCCGAGAAGACTGCACCTTCTGCTGGTCCAAAGTTTACG 2220  
Db 2716 CAGCCCTGCCAGATCCCTGCCAGATGACTGTCAATTGACCAAGCTGGTCCAAAGTTTCT 2775  
QY 2221 CCTGTCTCCAGGAATCTGTGAAGCCACAAAAGTAGGGCGGACAGCTCAAGGGAAAAAGC 2280  
Db 2776 TCATGCAATGGAGACTGTGTGTGAGTGTAGGACCAGAAAGCGCACTCTTGTGGAAGAAAGT 2835  
QY 2281 AGAAAGAGAGAGAAATGCCAGATTTCTGACTTTTACCCCTCTAGTGGAGACAGAACTATGT 2340  
Db 2836 AAAAAGAGGAAATGTAAAAATCCCATTTGTATCCCTGATTGAGACTCAGTATTGT 2895  
QY 2341 CTTGTGATGAAATTTATCCCACTTTATGGAACCTGTGCAATTCCTCCAGAA 2400  
Db 2896 CTTGTGACAAATATAATGCACAACTGTGGGAACTGGTCAGACTGTATTTTACCAGAG 2955  
QY 2401 GGCAGAGGGAGCCTCACCGAGGACTCGGGTACAGCAGACAGCAAGAAATGTGAGAA 2460  
Db 2956 GGAAGTGAAGTGTGTCTGGGAATGAAAGTACAGGAGACATCAAGGAATCGGACAA 3015  
QY 2461 GGCCTGCGCTTTTCAGCAGTAGCCTGTCTGTATAAAATGGAAGACCTGTGTGACCCCTCC 2520  
Db 3016 GGATATCGTTACCAAGCAATGGCATGTACGATCAAAATGGCAGGCTTGTGGAACATCT 3075  
QY 2521 TTCTGCGAGCTCTGCTTACATTCAGAAATATGTGTCAATCCCTGCCCATTTGATTGC 2580  
Db 3076 AGATGTACACCCATGTGTCAATTTAGGAGGCGCTGCATCATCCCTGCCCTCAGACTGC 3135  
QY 2581 AAGTTAAGCAGATTGCTAGTGTGGGGTCTTTCAGATTTCATCTTGTGGAATTTGAGTGAGA 2640  
Db 3136 AAGCTCAGTGTAGTGTCCAACTGCTCGCTCGCTGCAGCAAGTCTCTGTGGAGTGTGTGAAG 3195  
QY 2641 ATTGTATCCAAATGGCTTAAAGAAAAACCTTACATGGAGGAGACCAATGTCCTCAAACTG 2700  
Db 3196 GTTCGTTCTAAATGGCTGCGGTGAAAAACCATATAATGGAGGAAGGCGCTTGGCCCAAACTG 3255  
QY 2701 GATC-----TCAAGAAATCAGGTACATGAGCAGTCCCATGTTTACAGTGTAGTCCCAATCAG 2754  
Db 3256 GACCATGTCAACAGGACAGAGGTGTATGAGGTGTGCCATGTCCACAGTGTACTGCACACAG 3315  
QY 2755 TATTCCTGGGTGTAGAACACTGGTCTTTCATGCAAAATCAACAAATGAGCTGTAGGTCCTG 2814

Db 3316 TACCTATGGGTCAAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATCGGGAG 3375  
QY 2815 CGCTGTGAGAGGAAACACAATCTTAGGAATCATAGATGTGT---GAATACTGCGGATGGT 2871  
Db 3376 AACTGTGAGAGGGGTGCAAAACCCGAAAGTGAATGATGATGCAAGAAATACACAGAGATGGC 3435  
QY 2872 GAAGGTGAGCAGTGGATAGCAACCTGTGCAACAGGATGAAATTTCCCCAGAGAACCCAG 2931  
Db 3436 CTTCTGAACATGTAGAGGATTACTCTGTGACCAAGAGATGCCCCCTGGCTCTAGA 3495  
QY 2932 TCCCTGTTCTTATGTGTCGCCAATGAGTGTCTGTCTGTGAGTGGGACTTTGAGCAAA 2991  
Db 3496 GTGTGCAAAATTAACATGCCCCCTGAGACTGTGTGATATCTGAATGGGTGCCATGACCCAA 3555  
QY 2992 TGCCACAGTCTATGCGATPCCCCACACAAT---GAGAGAGAAATCCGCCACTCTCTAAGA 3048  
Db 3556 TGTGTTTTCCTTGCATCAAAAGCAGTTTCGGCAAGGTTCAGCTGATCCCATCAGACAA 3615  
QY 3049 CCATCACTGAATCAAGACTTTGTCTGAAGACTCAAGGTGAGCCCTTGCTCTCTGAAT 3108  
Db 3616 COAGCTGATGAAGAGATCTTGGCCCTAATGCTGTGAGAAAGAACCTGTAACTCTGAAC 3675  
QY 3109 GAAATTTCTTCCAGTTCAGTACAAATCTAACAGAGTGGAGCACATGCCAGCTCAGTGAA 3168  
Db 3676 AAAAATCTCTACCACTATGATTATACTAACAGCTGGAGTACATGTCACTGAGTGAG 3735  
QY 3169 AACGACCCCTGTGTCAAGGCTCAGGACCCGCTCTAAGCTGTGTGTGAGTGTGAGTGGC 3228  
Db 3736 AAGCAGTCTTGTGAAATGGAATAAAAACAAGATGTTGGATTTGTTCGAAGTGTGAGC 3795  
QY 3229 AAGCAGTCTCAGTGGACCAATGTGAGCAGCATAAATTTGGAGAGACCCAGAGAAATGAGC 3288  
Db 3796 AAGTCAGTTGACCTGAATATTTGTGAAGCCTTGGCTTGGAGAGAACTGGCAGATGAAC 3855  
QY 3289 ATTCCCTCTTGTGGAATGCGTCAACTGTCACTGTCACTCTCAGGGTGGAGCGCTTGGACA 3348  
Db 3856 ACGTCTGCATGTGGGAATGCCCTGTGAACCTGTGAGCTTTCTGATTTGCTCTCTTGGTCA 3915  
QY 3349 GAGTGTTCACAGACTGTGGCCATGAGGCTGATGAGCGGACTCCGATTTATCATTTG 3408  
Db 3916 GATGTTCTCAACATGTGGCTCAGAGAAATGATTCGAGACGAAACAGTGAACCCAG 3975  
QY 3409 CCAACCCAGAGAGAGGACGCGCATGCCCCACAGACTTTACCCAGGAGAAACCTGCCCA 3468  
Db 3976 CCGTTTCAAGGTGATGGAAGACCATGCGCTTCCCTGATGACCAAGTCCAAACCTGCCCA 4035  
QY 3469 GTGACCCCTCTACAGCTGGTCTTGGCAACTGGTCTGCATGATAATTTGGAGGTGGA 3528  
Db 4036 GTGAAGCCTTGTATCGGTGGCAATATGGCCAGTGTCTCCATGCCAAAGTGCAGAGGCC 4095  
QY 3529 GACTGTGGGGAAGAGATTACATCCGAGCCCTTTCTGCTGATGCTCCACAGTGGTTCAATA 3588  
Db 4096 CAGTGTGAGAGAGGAGCCAGAAACAAGGAACATTTCTTGTGTAGTAACTGATGGTCACT 4155  
QY 3589 TCTCATGAGCTGAGACGTGTGAGATGCATGTGTGGAGAAATGCCCTT----- 3638  
Db 4156 GATGATTTTCAAGAAAGTGTGGATGAGGAATTTCTGCTGCACTTGAACCTCATTTATGAT 4215  
QY 3639 -----TCAGGACAGCATCCTGAAGCAGCTGTGTCTGCTGCTGCCAGGAGACTGCCAT 3693  
Db 4216 GGTAAATAAATATGTTCTGGAGAAATCCCTGCAGCAGCCCTTGCAGGCTGACTGTAT 4275  
QY 3694 TTACAGAAATGGTTCAGAGTGGAGCAGATGTAATTAACCTGCAATGATGAGAGAGCTTT 3753  
Db 4276 TTGAAGGACTGTCTTCTCTGGAGCCTGTGTGAGCTGACCTGTGTGAATGTGTGAGGATCTA 4335  
QY 3754 GAGACTGTGGGCCCGCCAGTCTAGATCAAGGACTTTTAAATTTCACTTTTGAAGAACAA 3813  
Db 4336 GGCCTTGTGGAATACAGGTCAAGTCCAGACCGGTGATTATACAGAACTAGAGAAATCAG 4395  
QY 3814 GACAGCTGCCCCCAACAGGTTCTTAGAAACAGCCCTTGTACAGAGGCAATGTTATCAC 3873

4396 CATCTGTGCCAGCAGCATGTTTGAACAAATCATGTTATGATGACAGTCTATGAA 4455  
3874 TACACATGGAAGCAAGTCTTTTGGAAACAATAACGAACAGTCTATGTTGCCAGGTTCA 3933  
4456 TATAATGATGCCAGTCTTGGAGGGCTCTTCCGACAGTGTGGTCAAGGTCA 4515  
3934 GATGGCTTAATCTACAGGAGCTCTCCCTCAGGCGCTCTGCTGCCATTCGGCAG 3993  
4516 GATGGTATAATGTAACAGGGGCTGTTGGTATGAGCAGCCTGATCCGACAGTCT 4575  
3994 TGCATTCAGCTCGACAAACCTTTCTCTACTGTACACAGGTGGTGTGTGT 4053  
4576 TGTAAACCCCGTGTAGTACACCCCTCTGTCTGTAGCAGACAAACATGCCATGT 4635  
4054 GAGAAGGCTATACAGAGATAATGAATCAATGTTTCTCTGATTACTGCA----- 4105  
4636 GAAGAAGGGTACACTGAAGTCATGTTCTTAACAGCACCTTGGACATGCACATTATC 4695  
4106 -----TGAAAGTACCAGCTCAGAGGATAAAAGCTGATGTGAAAACCTTTCTG-G 4157  
4696 CCGTGGTGTATTAACCCATGAGGAGCAAAAGAGAGATGTGAAAACCAAGTGGCT 4755  
4158 GAAAGACAGACCTGTGAATTCAAAATAACATGATATTTTAAAGGA--TGCTCTTCAA 4215  
4756 GTACATCCAAACCAACCTCCAGTAACCCAGCAGGACGGGAGGACTGTTTCTACAG 4815  
4216 CCACTTATCCAGATGGCGAGTAAATTTGGTTTATGCGGTTTCAGGTGGCGTTT 4275  
4816 CCAATTTGGGCAAGTGGAGACTAAAGACCTGGTTTACGGTGTAGCAGCTGGGCAAT 4875  
4276 CTCATCATGATTTCTTAATATTACTTCTACTGTTTGAAGAGCA--AAACCA 4332  
4876 GTGTTACTCATCTTTATTTGTTCTCATGATTTATCTAGCTTGCAAAAGCAAGAACCC 4935  
4333 CATCAAGCACACTCCCAACAGAGCCCTGACCTTAGCTTACGATGAGAGACTTAGAC 4392  
4936 CAAAGAAGGCAAAACCAACCGACTGAAACCTTTAACTTAGCCTATGATGAGATGCCGAC 4995  
4393 ATGTAA 4398  
4996 ATGTAA 5001

RESULT 9  
AI58451  
ID AAI58451 standard; cDNA; 6378 BP.  
CX AAI58451;  
CX 22-OCT-2001 (first entry)  
CX Human polynucleotide SEQ ID NO 654.  
CX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
CX peripheral nervous system; neuropathy; central nervous system; CNS;  
CX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
CX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
CX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
CX leukaemia; ss.  
CX Homo sapiens.  
CX WO200153312-A1.  
CX 26-JUL-2001.  
CX 26-DEC-2000; 2000US-00534263.  
CX 23-DEC-1999; 99US-00471275.  
CX 21-JAN-2000; 2000US-00488725.  
CX 25-APR-2000; 2000US-00552317.  
CX 20-JUN-2000; 2000US-00598042.  
CX 19-JUL-2000; 2000US-00620312.

03-AUG-2000; 2000US-00653450.  
14-SEP-2000; 2000US-00662191.  
19-OCT-2000; 2000US-00693036.  
29-NOV-2000; 2000US-00727344.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX WPI; 2001-442253/47.  
DR P-PSDB; AAN39295.  
XX Novel nucleic acids and polypeptides, useful for treating disorders such  
as central nervous system injuries.  
PS Claim 1; SEQ ID NO 654; 10078pp; English.  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
CC encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX SQ Sequence 6378 BP; 1823 A; 1420 C; 1619 G; 1515 T; 0 U; 1 Other;  
Query Match 28.08; Score 1229.6; DB 4; Length 6378;  
Best Local Similarity 57.6%; Pred. NO. 0;  
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;  
QY 10 TGCATTTCAGAGCTGAAACCACTGTGTTTGCATATATGCGAATATGCGAACACTTTTGCCTT 69  
DB 541 TGCATCCAGAAAGCAAGACATTCCTGCGAGGATATCATCTGTGAGTACTTTGAGCCC 600  
QY 70 CAGCTCTCTACAGAACAGGCTTGGCTCATCTCTGTCCTCCCGGATTTGTAGTATCTGAG 129  
DB 601 AAGCCTCTCTGAGCAGGCTTGCCTCATCTCTTGCAGCAAGATTGCATCTGTCTGAA 660  
QY 130 TTCTTACATGTTCCAACTCTAGCAAGGATGTGGAGAAATTCAGCATAGAACTCGC 189  
DB 661 TTTTCTGCTGTTCCGATGCTCCAGACCTGCGGAGCGGCTCCAGCACCGGCGGT 720  
QY 190 GCGTCTATAGTCTCCCTCTCTTTGTTGTTTGCATATGTCCTCAATCTGATGAGTCAAGA 249  
DB 721 CATGTGTGGCGGCGCGGAGTTCGGAGGCTCTGGCTGTCTCAACCTGACGAGTTCAG 780  
QY 250 GCCTGTGATGCTCCCATTTCTCTCTCTTGGGAGAGGAATATACATTTAGCCTTAAG 309  
DB 781 GTGTG-----CMAATCCAGTCCATGCGAGCGCGAGGCTCAGGTACAGCCTGCAT 831  
QY 310 GTTGACCATGGAGTAAATGACAGCTGCTCATCTTTAAAGAAATTAATCAAGCGGAGA 369  
DB 832 GTGGGCGCTTGGAGCACCTGCTCAATGCCCCCTCCGACACAGTAAAGCAAGAGAGA 891  
QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTTCAATGAG 402  
DB 892 CGCGGGAAGATTAAGAACCGGAAAGGACCGCAGCAAGGATTAAGGATCCAGAGCC 951  
QY 403 CGAGTCACTTTTAAACATCAAGTGTACA-----AAGCATCATCATCTTCAAGTCT 453  
DB 952 CGCGAGCTTTTAAAGAAAGAAAGAACAGAACAGAGAACAGGACAGACAGACAAATAT 1011

454 TGGCAATAGAGATAGGTTATCAAAACCGGCGAGGTTTCGTGTACAGAAAGTGTAGACAA 513  
1012 TGGACATCCAGATTGGATATCAGACAGAGAGGTATGTGCAATTAACAAGACGGGAAA 1071  
514 AATGCTATGTTAAGCCTTTGGCTTCAAGATTCCCTCCATTGACTGTTCAAGTCTCTGATC 573  
1072 GCTGCTGATTTAAGTTTGGCCAGCAGAGAGCTTCAATGACCTTCCAGTCTCTGTGTG 1131  
574 ATGCCCAAAGACTGTGAAAACCTCCAGTGTGTCCTCTGGAGCCCTCTCTCCAAAGACATGC 633  
1132 ATCAACAAAGAGTCCAGGTTTCCGAGTGTGTCAGAGTGGAGCCCTCTCTCAAAAACATGC 1191  
634 CGTTCCAGGAGCTCTCTGCCAGGATTTAGAGAGCAGGAGCCGGACGTAAGACATCGCT 693  
1192 CATGACATGGTGTCCCTGCGAGGCACTGCTGTAGAGCACCAACCATCAGCGAGTTTCCC 1251  
694 ATTGGAGTGGAAAGGAGTGTCTGAACTTCTTGAGAAAGGCGCTGCTATTTGTAAGGA 753  
1252 ATTGGCAGTGAAGAGGAGTGTCCAGAAATTTGAAGAAAAGAACCTCTTTGTCTCAAGGA 1311  
754 GAACTTCTGCAGCAATGTCCAGGTATTCCTGGAGAACTTCTGNAATGGAAGAAATGCCAA 813  
1312 GATGGAGTTGTCCCTGTGCCACGTATGCTGTGGAGAACTACAGAGTGGACTGAGTGCCT 1371  
814 GTCTCTCTCTCCTCGAGCAGCAGGATCCCACTGGCATGTGTGACGGACCCGCTGTGTGGC 873  
1372 GTGGACCCCTTTGCTCAGTCAGCAGGACAAAGAGGCGGCGCAACAGACGGCCCTCTGTGGA 1431  
874 GGTGGATCCAGACCGGAGGTGTACTGTGCCAG-----AGGTAACAGCA 921  
1432 GGGGCATCCAGACCGGAGAGGTGTACTGGGTGAGGCCAAACGAAAACCTCTCTCAAA 1491  
922 GCTGCCGACCTGAGGGCCCAAGAAAGTCTCTAGACTGTGGAAGAGCATTTATGTGGGA 981  
1492 TTAAGTACCCACAGAACAAAGAGCTCAAGCCAAATGAAATTTATGCACTGGA 1551  
982 CCGCCCGGTTGCCCTCTCAGCTGTGCAATATCCCTTGCTCTACGGACTGTCAATATCT 1041  
1552 CCTATCCCTAATACTACACAGCTGTGCCACATCTCTGTGTCAACTGAATGTGAAGTTCA 1611  
1042 TCCTGGTCAGCTCGGGGCTGTGATCCATGAAACTGTATGAACCTCAGGGGMAAAA 1101  
1612 CTTTGGTCAGCTTGGGACCTGTACTATTATGAAAATGTATGATCAGCAGGMAAAA 1671  
1102 GGAATTAAGACAGGCGACG-----CAATGTCTCATGGAACTACAGGCGCTGCA 1152  
1672 GGCCTCAAACTGAGAAAGCGCGCATTAACCAATGAGCCACTGGAGGCTCTGGGTAAC 1731  
1153 GGGCATTTGCCCTCATTTGGTGGAGTGTCTTCTGTGAGGATCCAAATGTCTACCGATGG 1212  
1732 GGAATCCCTCTACTTACTGGAAGCCATTCCTCTGGAAGGCGCTGCTGTATGACTGG 1791  
1213 CTGGCA---TCAGAAAGGATCTGTTTCCCTGATCATGGA---AATGTGGCCTGGACAT 1266  
1792 AAAGCAGTGAGACTGGGAGACTGGAGCAGATACGGAAGAGGAGTGTGTTCCAGGCAAG 1851  
1267 CGTATCTGAAGGCGCTGTGCAGATGACCGCGGAGAGATGTATCAGGAGTCTTTGC 1326  
1852 CAAGTTCAAGAGGTTGTGTGATCAACAGATGATGAGAGAAGTTGACAGACAGCTGTGC 1911  
1327 CCAGTTCCCTCTCTCTGAGAGGAGTCTTGTGAAATTCCTGCCGAATGGACTGTGTG 1386  
1912 AGAGATGCCATCTTCCCATCTCCCTGTGGCTGTGATGCCCATGCCCCGAAGACTGTGTG 1971  
1387 CTGACGAGTGGAGGAGTGTCTCTGTTCCAGTCTGTTTCAATAAATAAATACTCAGAT 1446  
1972 CTCAGCAGATGCTTACGTGTCTCTCTGTCTCACACACCTGCTCAGGGAAAAGCAGACAA 2031  
1447 GGGAAACAGACAGGTCAAGAACTATCTGTGC-----ACTGGCTGGGAGGTGGAAGCCCA 1503  
2032 GGGAAACAGATACAGACAGCATCAATCTGTGCCCTATGCGGGTGAAGAGGTGGAATTCGC 2091  
1504 TGTCCCTCTAGTCAAGGCTCTCCAGAGCATCGTTTGTGTAAATGACCATCTCTGTATGACG 1563

2092 TGTCAAAATAGCAGTGTCTTGCAGAGAGTCAAGAGCTGTAAATGAGCATCTTGTGCAGGTG 2151  
1564 CTTTCACTGGGAGACATCGCTTGGGGCCCTTGTCTGAGGACACATTGTTAACTGCCCTT 1623  
2152 TACCACTGGCAACTGTGTTCCCTGGGGCCAGTGCATTGAGGACACCTCAGTATCTGCTTC 2211  
1624 AATCAACCAATTTGGCTGGAATGGAGAACCCAGTGTGTTAGGCAATTCAGACTCGGAGA 1683  
2212 AACCAACTTACGACTTGAATGGGAGGCTCTCTGCTGTGCGCATGCAGACAAGAAAA 2271  
1684 GTCCTCTGTGTCAAGAGTCAGTGGGACAAAGTAATGACCAAAAGATGCCAGATTCCTACT 1743  
2272 GTCACTGTGTGCGAGTCAATGTGGCCCAAGTGGGACCCAAAATATGCTCTGAAAGCCTT 2331  
1744 CGACTCTGAACTGTGCGCCCTCTGTTTCTCCCATGCAAAAAAGACTGTATTTGTGACTGCT 1803  
2332 CGACTCTGAACTGTGAAGCCTTGTCTGCTTCTTGTGAAGAGGACTGTATTTGTGACCCCA 2391  
1804 TTCAGTGTGTGACACCTGCCCCCAAGGATGTCCNAGCAGGAATGCCACAGT---AAAA 1860  
2392 TATAGTGTACTGGACATCATGCCCCCTCTTCTGTGAAGAGGGGACTCCAGTATCAGGAAG 2451  
1861 CAGTCTTCGATACAGAAATCATCATCAAGAGCAGCCAAATGGAGGCCAGAAATGCCAGAT 1920  
2452 CAGTCTAGGCATCGGCTCATCTTCTGCTGCGAGCCAAACGGGGCCGAGACTGCACAGAT 2511  
1921 ACCTTATATAGGAGAGAGTGTGAAGATGTTTCTTGTGTCTGTATATTCGTTGAG 1980  
2512 CCCCCTCTATGAAGAGAGGCTGTGAGGCACCTCAAGCGTCCAAAGCTACAGGTGGAAG 2571  
1981 CCACAGAAATGAGGCCCTTGCACTTCTAGTGCAGAGTCTGTCTGCGAGGGAATAACGGGC 2040  
2572 ACTCACAAATGGCGCAGATGCCAATTAAGTCTTCTGAGCGTGCACACAGACAGCCCTGA 2631  
2041 AGCAGTGAAGCCTGTGGAAGAGGGTTACAAACAGAGCTGTCTCATGCTATCTCTGATGAC 2100  
2632 GCACAGGAAGGCTGTGGGCTGGGCGACAGCAAGAGCCATTACTTGTGCGAAAGCAAGAT 2691  
2101 AACCCTCAGCAGAGAAATGATGAATGCTCTCAAGCAGACAAACGGCATGCCCTCTCTGTG 2160  
2692 GGAGACAGGCTGGAAATCCATGAGTGCCTCAGTATGACAGCCCTGTGCCAGCCCTTACC 2751  
2161 CAAGAATGCACAGTCCCATGTGGAAGACTGCACTTCACTGCTTGTGTTGTTCAAGTTACG 2220  
2752 CAGGCTGTCCAGATCCCTGCGAGGATGACTGTCAATGACAGCTGCTGCTCAAGTTTCT 2811  
2221 CCTGCTCCACGACTGTGAAGCCACAAAAGTAGGCGGCGACAGCTCACAGGGAAGAC 2280  
2812 TCATGCAATGGAGACTGTGTCAGTTAGGACCCAGAAAGCGCACTCTTGTGGAAAAAGT 2871  
2281 AGAAAGAGGAGAAATGCCAGGATTTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
2872 AAAAAAATAAATAATGTAATAATTTCCATTTGTATCCCTGATTGAGACTCAGTATGT 2931  
2341 CTTTGTGATGATTTATATCCCACTTATGGAAGTGTGGAAGTGTGATGCTTCTTCCAGAA 2400  
2932 CTTTGTGCAATATATTAATGCACAACCTGTGGGAACCTGCTCAGACTGTATTTTACCAGAG 2991  
2401 GGCAAGAGGAGCCCTCACCGAGGACTGGGCTGTAAGCAGACAGCAAAAGAAATGTGGAAA 2460  
2992 GGAAGTGGAGTCTTCTGTTGGAAATGAAGTACAGGAGACATCAAGAAATCGGACAA 3051  
2461 GGCCTGCGCTTTCGAGAGTACCTGTTCTGATAAAAATGGAAGACCTGTTGACCCCTCC 2520  
3052 GGATATCTGTACCAAGCAATGGCATGTACGATCAAAATGGCAGCTTGTGGAAACATCT 3111  
2521 TTTCTCAGCAGCTGTGTTACATTCAGAAAAATGTTGCTATTCCCTGCCCTTGTGATGTC 2580  
3112 AGATGTAACAGCCATGGTTACATTTGAGAGGCGCTGCAATCCCTGCTCCCTCAGACTGC 3171  
2581 AAGTTAAGCGATGTGTCTAGTTGGGGTCTTTCAGATTCATCTTGTGGAATTTGGAGTGA 2640





XX 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 19-JUL-2000; 2000US-00620312.  
XX (ZHOU/) ZHOU P.  
PA (TANG/) TANG Y T.  
PA (LIUC/) LIU C.  
PA (ASUNDI V. DRMANAC R T.  
XX (DRMA/) DRMANAC R T.  
XX Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;  
XX WPI; 2003-678194/64.  
XX  
XX New polynucleotide, useful for treating diseases e.g., cancer or  
XX neurodegenerative diseases.  
XX  
XX Claim 1; SEQ ID NO 332; 99pp; English.  
XX  
XX The invention relates to a polynucleotide comprising a sequence given in  
XX the specification, or its mature protein-coding portion, or its  
XX complement. The polynucleotide is useful for treating diseases e.g.,  
XX cancer or neurodegenerative diseases and many others listed in the  
XX specification. The present sequence represents a novel human cDNA. Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format directly from USPTO  
XX at [seqdata.uspto.gov/sequence.html?docID=20030104529](http://seqdata.uspto.gov/sequence.html?docID=20030104529).  
XX  
XX SQ Sequence 6378 BP; 1823 A; 1420 C; 1619 G; 1515 T; 0 U; 1 Other;  
Query Match 28.0%; Score 1229.6; DB 8; Length 6378;  
Best Local Similarity 57.6%; Pred. No. 0;  
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;  
10 TGCATTGAGAGCTGAACCGAAGCTGTGGTGGCAAAATGGAATATGCGAACACTTTGGCCCTT 69  
541 TGCATCCAGAAAGACAAAGACATTCCTGCGAGGATATCATCTGTGAGTACTTTGAGCC 600  
70 CAGCCTCCTACAGAACAGGCTGCTCATTCCTGTCCTCCCGGATGTTGATGATCTGAG 129  
601 AAGCCTCTCTGAGAGGCTGCTGCTCATTCCTGTCCTGAGCAAGATTGCATCGTGTCTGAA 660  
130 TTCTTACATGCTCCAACTGTAGCAAGGATGTGGGAAGAAATGCGACATAGAACTCGC 189  
661 TTTTCTGCTGCTCCGAATGCTCAAGACCTGCGGACGCGGCTCCAGCACCGGACGCGT 720  
190 GCGGTCTAGCTCCCTCTCTCTTGGTGGTGTGCAATGTCCAAATCTGACTGAGTCAAGA 249  
721 CATGTGTGGCGCCCGCGAGTTCGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCCAG 780  
250 GCCTGTGATGCTCCCATTTCTCTCTTGGGGAAGAGGAATATACATTTAGCCTTAAG 309  
781 GTGTG-----CCAATCCAGTCCATGCGAGCGCGAGGAGCTCAGGTACAGCCTGCAT 831  
310 GTTGGACCATGGAGTAATGAGACTCCCTCATCTTTAAAGAAATTAATCCAGCGGAAGA 369  
832 GTGGGGCCCTGGAGACCTGCTCAATGCCCACTCCCGACAAGTAAGACAAGCAAGGAGA 891  
370 ACTGTTCTGGATTTTAA-----CTGTGATTCAAATGAG 402  
892 CGCGGGAAGATAAAGAACGGGAAGACCGGACGAGTAAGGATCCAGAGCC 951  
403 CGAGTACCTTTTAAACATCAAGTTACA-----AAGCATCATCATTCGAAAGTCT 453  
952 CGCGAGCTTTTAAAGAAAAGAGAAAACAGAAAACAGGACAGACACAGAGAAACAATAT 1011  
454 TGGCAATAGAGATAGTTATCAACCCGCGAGTTCGTGTACAGAAATGATGACAA 513  
1012 TGGACATCCAGATTGATATCAGCCAGAGAGTTATGTGCATTTACAGACGGGAAA 1071  
514 AATGCTATGTTAAGCCCTTGGCTTTCAAGATTCCCTTCCATTGACTGTTCAGTCTCTGCATC 573  
1072 GCTGCTGATTTAAGCTTTTGCCACAAGAGAGCTTCCAATGACCTTCCAGTCTCTGTGTG 1131  
574 ATGCCCAAGAGCTGTGAAACCTCCAGTGGTCTCTCTGGAGCCCTGTCTCAAGACATGC 633  
1132 ATCACCRAAGAGTGCCAGGTTTCCGAGTGGTTCAGAGTGGAGCCCTGCTCAAAACATGC 1191  
634 CGTTCAGGAGTCTCTTCCAGAGTATTAGAGAGAGGAGCCGAAACGTGAAGACATGGCT 693  
1192 CATGACATGGTGTCCCTGCGAGGACCTGTTGTAAGGACACGAACCATCAGGACGTTTCCC 1251  
694 ATTGAGAGTGAAGAGGAGTGTCTGAACTTCTGAGAAAGAGGCTGCTGTTGTAAGGA 753  
1252 ATTGACAGTGAAGAGAGTGTCCAGATTTCAGAAAGAAAGAACCTGTTGTTCTCAGGA 1311  
754 GAACTTTCGAGCAATGTCCAGGTATTCCTGAGAACTTCTGAAATGGAAGAAATGCCAA 813  
1312 GATGAGTGTCTCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGGCGT 1371  
814 GTCTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGACGGGACCGTGTGTGGC 873  
1372 GTGGACCTTTGCTCAGTACAGAGACAAGAGGCGGGAACACAGAGCGCCCTCTGTGGA 1431  
874 GGTGGGATCCAGACCCCGGAGGTGTACTGTGCCAG-----AGCGTACCAGCA 921  
1432 GGGGSCATCCAGACCCGAGAGGTGTACTGCGTGAGCCCAACGAAACCTCTCTCAAA 1491  
922 GCTCCGCACTGAGGGCCCAAGAAAGTCTCTAGACCTGTGGAAGAGGATTTATGTGGGA 981  
1492 TTAAGTACCCCAAGAAACAAAGAGCCTCAAGCAATGGACTTTAAATATGACATGGA 1551  
982 CCGGCCCCGTTGCCCCCTCTCAGCTCTGCAATATCCCTTGTCTACGGACTGCATAGTATCT 1041  
1552 CCTATCCCTATACTACACAGCTGTGCCACATTCCTTGTCCAATGATGTAAGTTTCA 1611  
1042 TCCTGTGAGCCTCGGGCCTGTGCATCCATGAAACTGTGTATGAACTCAGGGGAAAAA 1101  
1612 CTTTGTGAGCTTGGGACCTTGTACTTATGAAACTGTATGATCAGCAGGGAAGAAAA 1671  
1102 GGATTGAAACGAGCGACG-----CCATGCTCTCATGGAATCTACAGGCGCTGCA 1152  
1672 GCTTCAAACTGAGAGAGCGGCGCATACCAATGAGCCCTGAGGCTCTGGGGTAACC 1731  
1153 GGGCATTGCCCTCAATTTGGTGGAGTCTGTTCTTGTGAGGATCCAAATGTCTACCGATGG 1212  
1732 GAAACTGCCCTCACTTACTGGAAGCCATTCCTGTGGAAGAGCCTGCTGTATGACTGG 1791  
1213 CTGGCA---TCAGAGGAGTCTGTTTCCCTGATCATGGAA---NATGTGGCTGGGACAT 1266  
1792 AAGCAGTGAAGTCTGGGAGACTGGAGCCAGATAACCGAAAGGAGTGTGTTCCAGGCAG 1851  
1267 CGTATTTCTGAAGGCGCTCTGCCAGAAATGACCGCGAGAAAGATGTATCAGGAGTCTTTGC 1326  
1852 CAAGTTCAAGAGGTTGTGTGCATCAACAGTGTGAGAGAAAGTTGACAGACACTGTGC 1911  
1327 CAGTTCCTCCCTCTCTGAGAGGAGTCTTGTGAAATTCCTCCCGAATGGAGTGTGTG 1386  
1912 AGAGATGCCATCTTCCCCCATCCCTGTGGCTGTGATGCCCATGCCGAAAGACTGTGTG 1971  
1387 CTGACGAGTGGAGCGGAGTGTGTCATCTGTTCCAGTCTCTTCAAAATAAAACTCAGAT 1446  
1972 CTCAGCATAAGTCTTACGTGTCTCTCTGCTCAACACCTGCTCAGGGAAACACAGAA 2031  
1447 GGGAAACAGACCAAGTCAAGAACTATCTGTGC-----ACTGGCTGGGGAAGTGGAAAGCCA 1503  
2032 GGGAAACAGATACAGACAGCATCCATTCGTGSCCTATGCGGCTGAAGAGGTGGAATTCGC 2091  
1504 TGTCCCTCTCTGAGGCTCTCCAGAGCATCTTGTGTGTAATGACCATTCCTGTATGCAG 1563  
2092 TGTCAAAATAGAGTCTTTTGCAGAAAGTACGAAGCTGTATGAGCATCTTGTGCACAGT 2151  
1564 CTTCACTGGGAGACATCGCTTGGGGCCCTTGTGTTCTGAGGACACATTTGGTAACCTGCCCTT 1623  
2152 TACCACTGGCAAACTGGTCCCTCGGGCCAGTGCATTTGAGGACACCTCAGTATCTGCTTTC 2211



Db 4372 GGCTTGGTGAATACAGTGCAGATCCAGACCGGTGATTTATACAGAACTAGAGATCAG 4431  
 3814 GACAGTGTCCCCCAACAGGTTCTTAGAAACACGCCCTTTGTACAGAGGCAAAATGTTATCAC 3873  
 Db 4432 CATCTGTGCCAGACAGATGTTTAGAAACAAAATCATGTTATGATGACAGTGTCTATGAA 4491  
 3874 TAGACATGGAAGCAAGTCTTTGGACAAATAACGAACGAACTATGTTGGTCCAGCTTCA 3933  
 Db 4492 TATTAATGGATGGCCAGTGTGTTGAAGGGCTCTCCGAAACAGTGTGGTCAAGGTCA 4551  
 3934 GATGGCGTTAATGTACAGAGGCTGTCCCTCAGCCCGCTCTGCTGCCAATCGGCAG 3993  
 Db 4552 GATGGTATAAATGTAAACAGGGGCTGTTGGTGATGAGCCAGCTGATGCCAGAGTCT 4611  
 3994 TGCATTCAGCCCTCGAAGAACCTTTCTCTACTGTATACACAGGCTGGAGTCTGTGGTTCT 4053  
 Db 4612 TGTAACCCACCGTGTAGTCAACCCACTCGTACTGTAGCGAGACAAACATGCCATTGT 4671  
 4054 GAGAAGGGCTATACAGAGATAAATGAATCAAAATGTTCTCTGATTACTGCA----- 4105  
 Db 4672 GAAGAAGGTACACTGAAGTCAATGCTTCTTAACAGACACCTTGAGCAATGCACACTATC 4731  
 4106 -----TGAAGTACAGGCTCAGAGATAAAGAGCTGATGTGAAAACCTTTCTG-G 4157  
 Db 4732 CCGGTGTGTATTACCACCATGGAGGACAAAGAGGAGATGTGAAAACCATCGGGCT 4791  
 4158 GAAAAACAGACCTGTGAATTCMAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
 Db 4792 GTACATCCAACCCACCTCCAGTAAACCCAGCAGGAGGAGGAGGAGGAGGAGGAGGAG 4851  
 4216 CCATTTGATCAGATGCGCGAGTAAATAATTTGGTTTATGGTTTACAGTGCGGCTTTT 4275  
 Db 4852 CCATTTGGCCAGATGGAGACTAAAGACCTGGTTTACGGTGTAGCAGTGCGGGCATTT 4911  
 4276 CTCATCATGATTTTCTTAATTTTACTTCTCTACCTGTTTGCAGAGGCA---AACA 4332  
 Db 4912 GTGTTACTCACTTTATGTTCTCCATGATTTATCTAGCTTGCAAAAGCCAAAGAACCC 4971  
 4333 CATCAAGCACACCTCCCAACAGAGCCCTCTGACCTTAGCTACGATGAGACTTAGAC 4392  
 Db 4972 CAAGAAGGCAAAACACCGACTGAACCTTTAACTTAGCTATGATGAGATGCCGAC 5031  
 4393 ATGTAA 4398  
 Db 5032 ATGTAA 5037

RESULT 11  
 AAI60237

AAI60237 standard; cDNA; 6292 BP.

AC AAI60237;

XX 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4226.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.

XX Homo sapiens.

OS WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US034263.

XX

PR 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00598042.  
 PR 19-JUL-2000; 2000US-00620312.  
 PR 03-AUG-2000; 2000US-00653450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX (HYSE-) HYSBQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 XX WPI; 2001-442253/47.  
 DR P-PSDB; AAM41081.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders such  
 as central nervous system injuries.  
 PT  
 PT  
 XX Claim 1; SEQ ID NO 4226; 10078pp; English.  
 PS  
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the  
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification

XX Sequence 6292 BP; 1818 A; 1391 C; 1595 G; 1487 T; 0 U; 1 Other;  
 SQ

Query Match

Best Local Similarity 27.1%; Score 1190.4; DB 4; Length 6292;

Matches 2594; Conservative 0; Mismatches 1786; Indels 132; Gaps 20;

QY 10 TGCATTGAGAGCTGAACCGAACTGTGTTGCAATGAAATATCGAACACTTTGCCCTT 69  
 Db 510 TGCATCCAGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 569  
 QY 70 CAGCTCTCTACAGACAGGCTTGCTCATTTCTTGTCCCGGGATTGTGTAGTATCTGAG 129  
 Db 570 AAGCTCTCTCTGGAGCAGGCTTGCTCATTTCTTGTCCCGAGCAAGATTGATCGTCTTGAA 629  
 QY 130 TTCTTTACCATGTTCCAACTGTAGCAAGGGATGTGGGAAGAAATTCGACGATAGAACTCGC 189  
 Db 630 TTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 689  
 QY 190 GGGGTATAGTCTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGACTGACAGA 249  
 Db 690 CATGTGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 749  
 QY 250 GCGTGTGATGTCCTCCATTTCTCTGCTCTCTTGGGGAAGAGGAATATACATTTAGCGCTTAAG 309  
 Db 750 GTGTG-----CCATCCAGTCCATGCGAGCGCGGAGGCTCAGGTACAGCTGCAT 800  
 QY 310 GTTGACCATGAGTAATGAGACTGCTGCTCATCTTAAAGAAATTAATCCAGCGGAGAGA 369  
 Db 801 GTGGGGCGCTGGAGCACCCTGCTCAATGCCCTCCTCCGACAAAGTAAGCAAGCAAGAGAGA 860  
 QY 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAAATGAG 402  
 Db 861 CCGGGGAAGATTAAGAACGGGAAAGACCGGAGCAAGAGTAAGAGTAAAGATCCAGAGGCC 920

403 CGAGTCACCTTTAAACATCAAAAGTTACA-----AAGCACATCATCTTCGAAGTCT 453  
421 CGCGAGCTTATTAAAGAAAGAGAAACAGAAACAGCAGCAAGACAGAGAACAAATAT 980  
454 TGGGCAATAGATAGTATTAACACCGGAGGTTCTGTGTACAGAGAGTATGAGCAA 513  
981 TGGGCAATCCAGATTGATATCAGACCCAGAGAGTTATGTGATTTAACAAGACGGGAAA 1040  
514 AATGCTATGTTAAGCCTTTGCTTCAAGATTCTTCCCATTCAGTGTTCAGTCTTCGATC 573  
1041 GCTGCTGATTAAAGCTTTTGGCCAGCAGAGAGAGCTTCCATGACCTTCCAGTCTGTGTG 1100  
574 ATGCCCAAGAGCTGTGAACCTCCAGAGTGTCTCTGTGAGCCCTGTGTCCAAAGACATGC 633  
1101 ATCAACAAAGAGTGCAGGTTTCCGAGTGTGAGAGTGGAGCCCTGTCTCAAAAACATGC 1160  
634 CGTTCAGGGAGTCTCTTGCAGAGATTAGAGCAGGAGCGGAAACGTGAAGCACATGGCT 693  
1161 CATGACATGGTCTCCCTCGAGCACTCGTGTAAAGACAGAACCATCAGGCACTTCCC 1220  
694 ATTGGAGTGGAAAGAGTGTCTGAACTTCTTGAAAGAGGCTGTGATTTGAAGGA 753  
1221 ATTGGCAGTGAAGAGGTCCAGAAATTGAAGAAAGAAAGAACCTGTGTCTCAAGGA 1280  
754 GAATCTCGACCAATGTCAGAGTATTCCTGGAGAACTTCTGAATGGAAGATGCCAA 813  
1281 GATGGAGTGTCCCTGTGCCAGCTATGCTGGAGAACTTACAGAGTGGAGTGGAGTGCCT 1340  
814 GTCTCTCTCCTCTCGAGCAGAGAGTCCCACTGGCATGTGACGGGACCCGTGTGTGGC 873  
1341 GTGGACCTTTGCTCAGTCAGCAGAGACAGAGCGCGGCAACAGACAGGCTCTGTGGA 1400  
874 GTGGAGTACAGACCCGGAGGTGTACTGTGCCAG-----AGGTACAGCA 921  
1401 GGGGCACTCCAGACCCGAGAGGTGTACTGCTGCGAGGCAACAGAAACCTCTCTCAAA 1460  
922 GCTGCGCACTCAGGCGCAAGGAAGTCTTAGACCTGTGGAAGAGCAATTATGTGGGA 981  
1461 TTAAGTACCACAGAACAGAGAGCTCAAGGCCATGGAATTAATATGACTGGA 1520  
982 CCGGCCCCCTGCTCAGCTGTGCAATATCCCTTGTCTTACGGAGTGCATATATCT 1041  
1521 CCTATCCCTAATACTACAGAGTGTGCCACATCTTGTGCCAACTGAATGTGAATTTCA 1580  
1042 TCTGTGTAGCCTGGGCTGTGCTATCCATGAATAACTGTGTAACCTCAGGGAAGAAA 1101  
1581 CCTGTGTAGCTTGGGACCTTGTACTTAAGAACTGTAAATGATCAGCAAGGHAAGAAA 1640  
1102 GGATTTAGAACGAGGACGG-----CCATGTCCTCATGGAATCTACAGGGCTGCA 1152  
1641 GGCTTCAAACTGAGGAAGCGCGCAATTAACCAATGAGCCCACTGGAGGCTCTGGGGTAACC 1700  
1153 GGGCATTTGCCCTCATTTGGTGGAGTGTCTCTGTGAGGATCCAAATGTCTACCGATGG 1212  
1701 GGAATGTCCTCTACTTACTGGAAGCCATTCCTGTGTGAGAGCCCTGCCTGTTATGACTGG 1760  
1213 CTGGCA---TCGAAGGGATCTGTTTCCCTGATCATGGA---AATGTGGCTGGGACAT 1266  
1761 AAAGCACTGAGACTGGGAGCTGCGAGCCAGATACAGGAAGAGTGTGTCCAGGACG 1820  
1267 CGTATTTGAAGCCCTGTCCAGAAATGACCGGAGAGAGATGTATCAGGGAGTCTTTGC 1326  
1821 CAAGTTCAAGAGGTTGTGTGCATCAACAGTGTGAGAGAGAGTTGACAGACAGCTGTGC 1880  
1327 CAGATTTCCCTCTCGAGAGGAGTCTGTGAAATTCCTCGCGGATGAGACTGTGTG 1386  
1881 AGAGATGCCATTTCCCAATCCCTGTGGCTGTGTGATGCCCCCATGCCCCGAAAGACTGTGTG 1940  
1387 CTGAGCAGTGGACGGAGTGTCTATCTTCCAGTCTCTGTTCAAAATAAAACCTCAGAT 1446  
1941 CTGAGCACATGTGTACTGTGTCTCTCTCTCTCAGACACCTGTGAGGGAAGACAGAA 2000

1447 GGGAAACAGACACAGGTCAAGAACTATCTCTGCC---ACTGCTGGGGAAGGTGGAAGCCA 1503  
2001 GGGAAACAGATACGAGCACCATCTCTGCCCCTATCGGGTGAAGAGGTGGAATTCG 2060  
1504 TGTCCCTCTAGTCAGGCTCTCCAAAGAGCATGTTTGTGTATGACCATCTCTGTATGCG 1563  
2061 TGTCCAAATAGCATGCTTTTGCAGAGAGTACGAAGCTGTATGAGCATCTCTGACAGTG 2120  
1564 CTTCACTGGGAGACATCGCTTGGGGCCCTTGTCTTGAGGACACATTGGTAATGCGCCTT 1623  
2121 TACCACCTGGCAACTGGTCCCTGGGCCAGTGCATTGAGGACACCTCAGTATCGTCTTC 2180  
1624 AATGCAACCAATTTGGCTGGAATGAGAGACCACTGTGTGTGTAGGCATTCAGATCGGAGA 1683  
2181 AACACAACTACGATTTGGAATGGGGAGGCTCTCTGTCTGTGCGGCATGCAACAAGAAA 2240  
1684 GTCTTCTGTCTCAAGAGTCACTGCGGACAGTAATGACCAAAAAGATGTCCAGATTTCTACT 1743  
2241 GTCATCTGTGTGAGTCAATGTGGGCCAAGTGGGACCCMAAAATGTCTCTGAAGCCTT 2300  
1744 CGACCTGAAAACCTGTGCGCCCTCTTTTCTCCATGCAAAAAGACCTGTATTTGACTGCT 1803  
2301 CGACCTGAAAACCTGTGAGGCTTGTCTGTCTTCTTGTAAAGAGACTGTATTGTGACCCA 2360  
1804 TTCAGTGAAGTGGACACCTGCCCAAGAT--GTGCCAAGCAGGAATGCCACAGT---AAA 1859  
2361 TATAGTACTGGAACATCATGCCCTCTTATCGTTAAAGAGGGGACTCCAGATTCAGGAA 2420  
1860 ACAGTCTCGATACAGAACTCATCCAAAGACAGCAATGAGAGGCCAGGAATGCCCAGA 1919  
2421 GCAGTCTAGGCATCGGTCTCATCTTTCAGCTGCCAGCAACCGGGGGCCGAGACTGCACA 2480  
1920 TACCTTATATGAGAGAGAGAGTGTGAAGATGTTCTTGTGTCTGTATATGCGTGG-- 1977  
2481 TCCCCCTATTAAGAGAGAGGCTGTGAGGCACCTCAAGCTGTGCCAAAGCTTACAGTGGGA 2540  
1978 -AAGCCACAGAAATGAGCCCTTGTGATCTTAGTGCC--AGAGTCTGTCTGGCAGGGAATA 2034  
2541 AGACTCAAAATGGGCACAGATGCCAAATTAGTCTCTTTGGAGCTGTGCACAGACAGCC 2600  
2035 ACGGCAGCAGTGAAGCTGTGTGAAAGGGGTTACAAACAGAGCTGTCTCATGATCTCT 2094  
2601 CTTGGAGCACAGGAAGCTGTGGGCTGTGGCGCAGCAGCAAGAGCCATTACTTGTGCAAG 2660  
2095 GATGACAAACCGGTGACAGAAATGATGGAATGCTCAAGCAGACAAACGGCATGCTCTC 2154  
2661 CAAGTGGAGGACAGGCTGGAATCCATGAGTGTCTACAGTATGACGGCCCTGTGCCAGCC 2720  
2155 CTTGTGCAAGAATGCAAGTCCCATGTTCGAGAAAGACTGCACCTTCACTGTCTGGTCCAAG 2214  
2721 CTTTACCAGGCTGCCAGATCCCTCTGCCAGATGACTGTCAATTGACCAGCTGCTCCAAG 2780  
2215 TTTACGCCCTGCTCCAGAACTGTGAAGCCACAAAAGTAGGCGGCGCAGCTCAGGG 2274  
2781 TTTTCTTCATGCAATGAGACTGTGTGAGTTAGGACCAAGAACGCGCACTCTTGTGGA 2840  
2275 AAAAGCAAGAAAGAGAAATGCCAGGATTTGACCTTTTACCCTCTAGTGGAGACAGAA 2334  
2841 AAAAGTAAAGAGGAATAATGTAAATTTCCATTTGTATCCCTGATTGAGACTCAG 2900  
2335 CTATGCTCTTGTGATGAATTTATATCCCAACCTTATGGAACCTGGTCAAGATTGCTTCT 2394  
2901 TATTGTCTTGTGACAAATAATAATGACCAACCTGTGGGAACTGGTCAAGCTGTATTTTA 2960  
2395 CCAGAAGGCAAGAGGAGCCCTCACCAGGACTCGGGTACAAGCAGACAGCAAGAAATGT 2454  
2961 CCAGAAGAAAGTGAAGTGTGTGCGGAATGAAGTACAAGAGACATCAAGGAATGC 3020  
2455 GGAGAAGGCTGCGCTTTTCAGCAGTAGCTGTCTCTGATAAAAATGAAAGACCTGTGTAC 2514  
3021 GGACAGGATATGTTTACCAGCAATGGCAATGTACGATCAAAATGGCAGGCTTGTGGAA 3080  
2515 CCTCTCTTCTGACAGCAGCTCTGTTGTATCTCAAGAAATGTGTCTATTCCTGCCCCATTT 2574







1408	GGCTTCAAAC	CTGAGGAACGGCGGCAATTACCAATGAGGCCCACTGGAGGCTCTGGGGTAACC	1467
1153	GGGCAATG	CCCTCATTTGGTGGAGTCTGTTCCTTGTGAGGATCCAATGTGCTACCGATGG	1212
1468	GGAAACTG	CCCTCATTCTACTGGAAGCAATTCCTGTGAAGAGCTGCTGTATGACTGG	1527
1213	CTGGCA---	TCAGAAGGATCTGTTTCCCTGATCATGGAA--AAATGGCCCTGGGCAT	1266
1528	AAAGCATG	TGAGACTGGGAACCTGGAGCCAGATAACGGAAGGAGTGTGGTCAGGCACG	1587
1267	CGTATTTCT	GAAGGCGCTCTCCAGAAATGACCGGGAGAAAGATGTATCAGGGAAGCTTTTGC	1326
1588	CAAGTTCA	GAGAGTTGTGTGATCAACAGTAGTGAGAGAAGTTGACACAGCTGTGC	1647
1327	CCAGTTC	CCCCCTCCTCTCTGAGAGAAAGTCTTTGTGAATTCCTTCCGAATGAGCTGTGTG	1386
1648	AGAGATGCC	ATCTTCCCATCCCTGTGGCCTGTGATGCCCGCTGCCCGAAGAACTGTGTG	1707
1387	CTGAGCGAG	TGGACGAGTGGTCAATCCTGTTCCAGTCCCTGTTCAATAAAAACTCAGAT	1446
1708	CTCAGCAC	ATGGTCTACGTGGTCTCTGCTCACACACTGCTCAGGGAACGACAGAA	1767
1447	GGGAAAC	CAGACCCAGGTCAAGAACTATCCTGGC---ACTGGCTGGGAAGGTGAAAGCCA	1503
1768	GGGAAAC	CAGATATCGAGCACGATCCATCTCTGGCCTATGCGGGTGAAAGGTGGAATTCGC	1827
1504	TGTCCTCC	CTAGTCAGGCTCTCCAAAGACATCGTTTGTGTAATGACCAATTCCTGTATGCAG	1563
1828	TGTCCTCA	ATATGACGTGCTTTGCAAGAAGTACGAAGCTGTGAATGAGCATCCTTGACAGTG	1887
1564	CTTCACT	TGGAGACATCCGCTTGGGGCCCTTGTTCGAGGACACATTTGGTAACTGCCCTT	1623
1888	TACCACT	TGGCAAACTGGTCTCCTGGGGCAGTGCATGAGGACACCTCAGTATCGTCTTC	1947
1624	AAATGCA	ACCATTGGCTGGAATGAGAGCCACGTTGGTGTAGGCATTACAGCTCGGAGA	1683
1948	AACACAA	CTTACGCTTGGAAATGGGAGGCTCTCTGCTGTGGCATGACAGACAAGAAA	2007
1684	GTCTCTCT	GTCTCAAGAGTCACTGGGACAAAGTAATGACCAAAAAGATGTCACAGATTTCTACT	1743
2008	GTCACTGT	GTGGAGTCAATGTGGGCCAAGTGGGACCCCAAAAATGTCTGAAAGCCTT	2067
1744	CGACCTGA	AACTGTGGCCCTGTTTCTCCCATGCAAAAAGACTGTATTTGTGACTGCT	1803
2068	CGACCTGA	AACTGTGAAGCCCTGTGCTCTTCTTGTGAAGAGGAGTGTATTTGTGACCCCA	2127
1804	TTCACTGAG	TGGACACCCCTGCCCAAGGATGTGCCAAGCAGAAATGCCACAGT---AAAA	1860
2128	TATAGTAC	TGGACATATGCCCTCTTGTGTAAAGAAGGGACTCCAGTATCAGGAAG	2187
1861	CAGTCTCG	ATACAGAAATCATCATCCAGAAGCAGCCAAATGGAGCCAGGAATGCCAGAT	1920
2188	CAGTCTAG	GCATCGGGTCACTTACGCTGCAGCCAAACGGGGCCCGAGACTGCACAGAT	2247
1921	ACCTTATAT	GAGGAGAGAGTGTGAAGATGTTTCCTTGTGTCTCTGTATATCCGTGGAAG	1980
2248	CCCCCTCT	ATGAAGAAGGCTGTGAGGACCTCAAGCGTGCACAAAGCTACAGTGGGAAG	2307
1981	CCACAGAA	ATGGAGCCCTTGCACTTAGTGTGCCAGAGTCTGTCTGGCAGGGAATAACGGGC	2040
2308	ACTCACAA	TGGCCAGATGCCAATTAGTCCCTTGGAGCGTGCACAAAGACAGCCCTTGA	2367
2041	AGCAGTGA	AGCCTGTGGAAGGGGTTTCAAAACAGAGCTGTCTCATGCATCTCTGTATGAC	2100
2368	GCA CAGGA	AGGCTGTGGGCTGCGGCGACAGGCAAGAGCCATTACTTGTGCAAGACAGAT	2427
2101	AACCGG	GT CAGCAAAATGATGGAAATGCTCAAGCAGACAAAACGSCATGCTCTCTTTGTG	2160
2428	GGAGGAC	CAGGCTGGAAATCCATGAGTGCTACAGTATGACGGCCCTGTGCGAGCCCTTACC	2487
2161	CAAGAA	TGCAAGTCCCATGTGCGAGAGAAGCTGCACCTTCACTGCTGGTCCAAAGTTTACG	2220
2488	CAGGCTGC	CAGATATCCCTGTGCGAGGATGACTGTCAATTTGACCAAGTGGTCCAAAGTTTCT	2547

2221	CCCTGCTCCACGAACGTGTGAAGCCACAAAAAGTAGCGGGCGACAGCTCACAGAGAAAAGC	2288
2548	TCATGCAATTGGAGACTGTGGTGCAGTTAGGACAGAGAAAGCGCACTCTGTTGGAAAAGT	2607
2281	AGAAAGAGGAGAAATGCCAGGATTCGTACCTTTACCCCTCTAGTGGAGACAGAACTATGT	2340
2608	AAAAAGAGGAAAATGTAAATAATCCCAATTTGTATCCCTCATTTGAGACTCAGTATTGT	2667
2341	CCTTGTGATGAATTTATATCCCAACCTTATGAAACTGTCAGATTGCATTTCTTCAGAA	2400
2668	CCTTGTGACAAATATAAATGCACAACCTGTGGGAACTGTCAGACTGTATTTTACCGAG	2727
2401	GGCAGAAAGGAGCCTCACCGAGGACTCGCGGTACAAGCAGACAGCAAAAGAAATGTGAGAA	2460
2728	GGAAAAGTGGAAAGTGTGTGGGAATGAAAGTACAAGGAGACATCAAGGAATCGCGACAA	2787
2461	GGCCTGCGCTTTTCGAGCAGTAGCTCTTCGTATATAAATGGAGAGCCTGTTGACCCCTCC	2520
2788	GGATATCGTTTACAGACAAATGCAATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCT	2847
2521	TTCTGCAGCAGCTCTGTGTTACATTTCAAGAAAAATGTGTCAATTCCTCGCCCATTTGATTGC	2580
2848	AGATGTAACAGCCATGTTTACATTTAGGAGGCGCTGCATCATCCCTGCGCCCTCAGACTGC	2907
2581	AGATTAAAGCATTTGTCATGTTGGGGTCTTCAGCTTCATCTGTGGATTTGGAGTGAGA	2640
2908	AACTCAGTGTAGTGTCTCAACTGTGTGCGCTCGAGCAAGTCTCTGTGGGAGTGTGTGAAG	2967
2641	ATTTCGATCCAAATGGCTAAAAGAAAACCTTACAATGGAGGACGACCATGTGCCAAACTG	2700
2968	GTTTCGTTCTAAATGGCTGGTGAAMAAACCATATAATGGAGGAAGGCGCTTGCCCAAACTG	3027
2701	GATC-----TCAAGAAATCAGGTACATGAGGCGAGTCCCATGTTTACAGTGCATACAG	2754
3028	GACCATGTCAAACAGGACAGGTGTATGAGTTGTCCCATGGCCACAGTCACTGCACACCAG	3087
2755	TATTCCTGGGTTGTAGAACACTGGTCTTCATGCAAAATCAAAATGAGTGCAGTGCCTTG	2814
3088	TACCTATGGGTACACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAG	3147
2815	CGCTGTGGAGGAGGAACAAACTCTAGGAAATCAGATGTT--GAATACTCGGATGGT	2871
3148	AACTGTGGAGAGGCGGTGAAACCCGAAAAGTGTGATGTCATGCAGAATACAGCATGGC	3207
2872	GAAGTGGAGCAGTGGATAGCAACCTGTGCAACCGAGGATGAAATTCCTCCCAAGAACCCAG	2931
3208	CCCTTCTGAAACATGTAGAGGATTAACCTCTGTGACCCAGAAAGATGCCCCCTGGGCTTACA	3267
2932	TCTCTGTTCTTTATGTGCCCAATGAGTGTGTCATGTCGTAGTGGGAGCTTTTGGAGCAA	2991
3268	GTGTGCAAAATTACCATGCCCTGAGGACTGTGTGTATATCTGAATGGGTTCAATGGACCCAA	3327
2992	TGCCCACTGATATGCGATCCCAACAAT--GCAGAGAAAGAACTGCGCACTCTGTCTAAGA	3048
3328	TGTGTTTTCCTTGTCAATCAAAAGCAGTTTCGGGCAAGGTCAGCTGATCCCATCAACACAA	3387
3049	CCATCACTGAATCTCAAGGACTTGTGCTGAAGACTCACAGTGCAGCTTGCCTCTCTGAAT	3108
3388	CCAGCTGATGAAGGAAGATCTTGCCCTAATGCTGTGTGAAGAAGAACCTGTAACTGAAC	3447
3109	GAAAATGCTTCAGTTTCAGTACAATCTAAACAGAGTGGAGCACATGCCAGCTGATGAAGAA	3168
3448	AAAAACTGTCTACCATATGATTTATATGTAACAGAGTGGAGTACATGTCACTGTGAGTGAG	3507
3169	AACGCACCTGTGGTCAAGGCGTCAGGACCGCGCTCTAAGCTGTGTGTGTCAGTGAATGCC	3228
3508	AAGGCAATTTGTGGAATTGGAAATAAAACAAGGATGTTGGATTTGTGTTTCAAGTGAATGCC	3567
3229	AAGCCAGTCAGCATGGAACCAATGTGAGCAGCAATAATTGGAGAGGCCCCAGAGAAATGAGC	3288
3568	AACTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAAAGAACTCTGGCAGATGAAC	3627



QY 70 CAGCCTCTACAGAACAGGCTTGCCCTCAATTCCTTGTCCTCCCGGATTTGTAGTATCTGAG 129  
Db 349 AAGCCTCTCTGGAGCAGGCTTGCCCTCAATTCCTTGCCAGCAAGATTGCATCGTGTCTGAA 408  
QY 130 TTCTTACATGGTCCCACTCTAGCAAGGATGTGGGAAGAAATTCAGCAGATAGAACTCGC 189  
Db 409 TTTTCTGCTGGTCCGAATGCTCCAAGACCTGCGGACGCGGCTCCAGCACCGGACGCGT 468  
QY 190 GCGGTCAATAGTCCCTCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGAGTCAAGA 249  
Db 469 CATGTGGTGGCCCGCCGAGTCTCGGAGGCTCTGGCTGTCTCAAACCTGACGGAGTTCCAG 528  
QY 250 GCCTGTGATGCTCCCAATTCCTGTCTCTTGGGAAGAGGAATATACATTTAGCCTTAAG 309  
Db 529 GTGTG-----CCTATCCAGTCCATGCGAGGCCGAGGCTCAGGTACAGCCTGCAT 579  
QY 310 GTTGGACCATGGAGTAATGAGACTGCTCATCTTAAGAAATTAATCCAAAGGGAAGA 369  
Db 580 GTGGGGCCCTGGAGCAGCTGCTCAATGCCCACTCCCGACAAGTAAGCAAGCAAGGAGA 639  
QY 370 ACTGTCTCGGATTTTAA-----CTCTGATTCAAATGAG 402  
Db 640 CGCGGAAGATAAAGAACGGGAAGAGCCGACAGCAAGAGGTAAAGGATCAGAGCC 699  
QY 403 CGAGTCACTTTAAACATCAAAAGTTACA-----AAGCACATCATCAATTCGAAGTCT 453  
Db 700 CGCGAGCTTATTAAGMAAAAGAGAAACAGAAACAGGACAGAACAGCAAGAGAACAAATAT 759  
QY 454 TGGCAATAGAGATAGTTATCAAAACCGCAGCTTTCGTGTACAAGATGATGGACAA 513  
Db 760 TGGGACATCCAGATTGGATAACAGACAGAGAGTTATGTGCATTAACAAGACGGGGAAA 819  
QY 514 AATGCTATGTTAAGCCTTTGCTTCAAGATTCTTCCCAATTCAGTGTTCAGTCTCTGCATC 573  
Db 820 GCTGTGATTTAAGCTTTTCCAGCAAGAGAAGCTTCCATGACCTTCCAGTCTCTGTGTG 879  
QY 574 ATGCCAAAGACTGTGAACCTCCAGTGTCTCTCGAGGCCCTCTGCTCAAGACATGC 633  
Db 880 ATCAACAAAGAGTCCAGGTTTCCAGTGTGTGACAGTGGAGCCCTCTGCTCAAAACATGC 939  
QY 634 CGTTCAGGAGTCTCTCCAGGATTTAGGAGCAGGAGCCCGGAACGTGAAGCACATGGCT 693  
Db 940 CATGACATGGTGTCCCTGTCAGGCACTCGTGTAGGACACGAACCATCAGGAGTTTCCC 999  
QY 694 ATTGAGTGTGAAGAGTGTCTGAACTTCTTGAGAAAGAGGCTCGATTTGTGAAGGA 753  
Db 1000 ATTGGCAGTGAAGAGAGTGTCCAGATTTGAAGAAAAGAACCTGTTGTCTCAAGGA 1059  
QY 754 GAACTTCTGCAGCAATGTCCTCAGGTATTCCTGGAGAACTTCTGAATGGAAGATGCAA 813  
Db 1060 GATGGAGTTGCTCCCTGTGCCACGTAATGGCTTGGAGAACTACAGAGTGAATGAGTGCCT 1119  
QY 814 GTCTCTCTCTCTCGAGCAGCAGGATCCCACTGGCATGTGACGGGACCCGTGTGGC 873  
Db 1120 GTGGACCTTTGCTCAGTCAGCAGCAAGAGGCGCGCAACCAAGACGCGCCCTCTGTGA 1179  
QY 874 GGTGGATCCAGACCCGGAGGTGTACTGTGCCAG-----AGGTACCAAGCA 921  
Db 1180 GGGGCGATCCAGACCCGAGAGGTGTACTGCGTGCAGGCCCAACGAAACCTCTCTCAAA 1239  
QY 922 GCTGCGCACTGAGGGCCAAAGAACTCTAGACCTGTGGAAGAGGATTTATGTGTGGGA 981  
Db 1240 TTAAGTACCCACAGAAACAAAGAGSCCTCAAAGCAATGGACTTAAATATATGCACTGGA 1299  
QY 982 CCGGCGCGTGGCCCTCTCAGTCTGGAATATCCCTTGTCTCTACGACTGATAGTATCT 1041  
Db 1300 CCTATCCCTTAATACACAGCTGTGGCAGATTCCTTGTCCAACGTGAATGTGAAGTTCA 1359  
QY 1042 TCTGCTCAGCCTGGGGCTGTGCATCAATGAATACTGTCAATGAACCTCAGGCGGAAAAA 1401  
Db 1360 CTTGGTCTGAGCTTGGGACCTTGACTTATGAAACTGTATGATGATCAGCAAGGAAAAA 1419

QY 1102 GGATTTAGAACGAGCGAGCGCC-----ATGTCTCATGGAATCTACAGGCGCTGCA 1152  
Db 1420 GCTTTCAAACTGAGGAAGCGCGCATTAACCAATGAGCCCACTGAGGCTCTGGGTAACC 1479  
QY 1153 GGGCAATGCCCTCAATTTGGTGGAGTCTGTCTCTTGTGAGGATCCAAATGTCTCCGATGG 1212  
Db 1480 GGAACCTGCCCTCACTTACTGGAAGCAATTCCTCTGTAAGAGCCTGCTGTATTGACTGG 1539  
QY 1213 ---CTGGCATCAGAAGGAGTCTGTTCCCTCATATGGA---AATGTGGCCTGGGACAT 1266  
Db 1540 AAGCGGTGAGACTGGGAGACTGGAGCCAGTAAACGGAAGGAGTGTGTTCAGGCAAG 1599  
QY 1267 GATATTCTGAAGCGCTCTGCCAGATGACCGCGAAGATGTATCAGGAGTCTTTGC 1336  
Db 1600 CAAATTCAAGAGGTGTGTGCATCAACAGTGTGAGAGAAGTGTGACACAGCTGTGC 1659  
QY 1327 CCAGTCCCTCCCTCTCCCTGAGGAGTCTTGTGAATTCCTCTCCGAAATGCAATGTGTG 1386  
Db 1660 AGAGATGCCATCTTCCCATCCCTGTGSCCTGTATGCCCCCATGCCCGAAGACTGTGTG 1719  
QY 1387 CTGAGCGAGTGGACGAGTGTCTCTGTTCCAGTCTCTTCAAAATAAAATCTAGAT 1446  
Db 1720 CTCAGCACATGCTCTACGTGTCTCTCTCTCTCACACACTCTCAGGGAACACAGAA 1779  
QY 1447 GGGAAACAGACAGTCAAGAACTATCTCTGSC---ACTGGCTGGGAGGTGGAAGCCA 1503  
Db 1780 GGGAAACAGATACGAGCAGATCCATCTGSCCTATGCGGCTGAGAGGTGGAAATTCGC 1839  
QY 1504 TGTCTCCCTTAGTCAGCTCTCCAGAGCATGTTGTGTATATGACCATTCCTGTATGAG 1563  
Db 1840 TGTCAAAATAGCAGTGTCTTCNAGAAGTACGAAGCTGTATGAGCATCTCTGCACAGT 1899  
QY 1564 CTTCACTGGAGACATCCGCTTTGGGCGCTTGTCTTGAGGACATTTGGTAATGCGCTT 1623  
Db 1900 TACCCTGCAAACTGGTCCCTGGGCGCAGTGCAATTGAGGACACCTCAGTATCGTCTTC 1959  
QY 1624 AATGCAACATTTGGTGGAAATGGAAGCCAGCTGTGTGTAGGCAATTCAGACTCGAGA 1683  
Db 1960 AACAACAATCAAGATTGGAATGGGAGGCTCTCTCTGTCTGCGCATGCAACAAGAAA 2019  
QY 1684 GTCTTCTGTCTCAAGATCAAGTGGGACAAAGTAATGACAAAGATGTCAGATTTCTACT 1743  
Db 2020 GTCACTGTGTGCGAGTCAATGTGGGCAAGTGGGACCCCAAAATGTCCTGAAAGCCTT 2079  
QY 1744 CGACCTGAACTGTGCGCCCTCTCTCCATGCAAAAGAGCTGATTTGTGACTGCT 1803  
Db 2080 CGACCTGAACTGTAGGCTTGTCTGCTTCTTGTGAAGAGACTGTATTGTGACCCA 2139  
QY 1804 TTCAGTGTGACACCCCTGCGCAAGGATGTGCCAAGCAGGAAATGCCAGT---AAAA 1860  
Db 2140 TATAGTACTGGACATCATGCCCTCTCTGTTAAAGAGGGACTCCAGTATCAGGAAG 2199  
QY 1861 GAGTCTCGATACAGATCATCATCAAGAGCCATGGAGCCAGTGGAGCCAGATGCCAGAT 1920  
Db 2200 CAGTCTTAGCATCGGCTCATCTTCACTGCGAGCACTCAAGCGTGCAGAACTACAGTGAAG 2259  
QY 1921 ACCTTATATGAGAGAGAGTGTGAAGATGTTTCTTGTCTCTGTATATCGGTGGAAG 1980  
Db 2260 CCCCCTATGAGAGAGCGCTGTGAGCACTCAAGCGTGCAGAACTACAGTGAAG 2319  
QY 1981 CCACAGAAATGAGCCCTTGCATCTTAGTCCAGAGTCTGTCTGCGAGGAAATACCGGC 2040  
Db 2320 ACTCAAAATGGCGAGATGCCAATTAGTCCCTTGGAGCGTGCAACAGACAGCCCTGGA 2379  
QY 2041 AGCAGTGAAGCTGTGGAAGGGGTTACAAACAGAGCTGTCTCATGCTCTCTCATGAC 2100  
Db 2380 GCACAGAAAGGTGTGGGCTGTGGCGACAGCAAGGCCATTAATTTGCGAAGCAGAT 2439  
QY 2101 AACCGCTCAGCAAAATGATGGAATGCTCTCAAGCAGACAAACGGCATGCTCTCTGTG 2160  
Db 2440 GGAGGACAGCTGGAATCCATGAGTGCCTTACAGTATGCAAGGCCCTGTGCGAGCCCTTACC 2499  
QY 2161 CAAGATGACAGTCCCATGTCGAGAGACTGCACCTTCACTGCTTGTGTCGAAGTTTACG 2220

2500 CAGGCTGCAGATCCCTGCGCAGGATGACTGCAATTGACAGCTGTCACAAAGTTTCT 2559  
2221 CCTGCTCCACGACTGTGGAAGCCACAAAGATAGCGCGACAGCTCACAGGGAAGAC 2280  
2560 TCATGCAATGGAGACTGTGTGAGTGTAGGACAGAAAGCCACTCTTTGTTGGAAGAGT 2619  
2281 AGAAGAAGGAGAAATGCCAGGATCTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
2620 AAAAGAAGGAAATGTAATAATTCCTATTTGATCCCTGATTGAGACTCAGTATTGT 2679  
2341 CCTGTGATGAATTTATATCCCAACCTTATGGAACCTGCTGATGATGATGATGATGAT 2400  
2680 CCTGTGACAAATATATACCAACCTGTGGGAACTGTGCTAGACTGATTTTACACAG 2739  
2401 GGCAGAAAGGAGCCTCACCGAGACTGCGGTCACAGCAGACAGCAAGAAATGTGGAGAA 2460  
2740 GGAAGATGGAAGTGTGCTGGGATGAAGTACAGGAGACATCAAGGAATGCGACAA 2799  
2461 GGCCTGCGCTTTGAGCAGTACGCTGTTCTGTATAAATGGAAGACCTGTTGACCCCTCC 2520  
2800 GGATATCGTTTACCAAGCAATGGCATGCTACGATCAAAATGGCAGGCTTGTGGAAACATCT 2859  
2521 TTCTGACAGACTCTGCTTACATTCAGAAATATGTCATTCCTGCGCATTTGATTC 2580  
2860 AGATGTAACAGCCATGTTTACATTTAGGAGGCTGCATCATCCCTGCCCTCAGACTGC 2919  
2581 AAGTTAAGCGATGCTAGTTTGGGGCTTTCAGTTTCACTTTGTGGAATGAGTGAGA 2640  
2920 AAGCTCAG\*GAGTGTGCTCAACTGCTGCGCTGCGCAAGTCTGTGGGAGTGGTGAAG 2979  
2641 ATTGATCCAAATGCTTAAAGAAACCTTCAATGGAGGACGACCATGTCCTCAAACTG 2700  
2980 GTTCGTTCTAAATGCTGCGTGAAACCATATATATGGAGGAGGCTTGTGCCCAACTG 3039  
2701 GATC-----TCAAGAAATCAGGTACATGAGGAGTCCCATGTTACAGTGAGTGCATCAG 2754  
3040 GACCATGTCAACAGGACACAGGTGTATGAGTGTGCTCCATGCCACAGTGCATCCACCAG 3099  
2755 TATTCCTGGTGTAGAACACTGCTTCTCATGCAAAATCAAAATGAGTGCATGAGTGCCTG 2814  
3100 TACCTATGGGTACAGAGCCCTGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAG 3159  
2815 CGTGTGGAGGAGAACACAATTCAGAAATCAGATGTGT---GAATCTCGGATGT 2871  
3160 AACTGTGGAGGCGCTGAAACCCGAAAGTGAATGATGCATGCAGATACACAGATGCC 3219  
2872 GAAGGTGGAGCAGTGATAGCAACCTGTGCAACAGGATGAATTCCTCCAGAAACCCAG 2931  
3220 CCTTCTGAAACATGTAGAGATTACCTCTGTGACCCAGAAAGATGCCCCCTGGCTCTAGA 3279  
2932 TCCTGTTCTTTATGTTCCCAATGATGTGTCATGCTGAGTGGGACTTTGGAGCAAA 2991  
3280 GTGTGCAAAATTACCTGCTGAGGACTGTGTGATATCTGAATGGGGTCCATGGACCCAA 3339  
2992 TGCCACAGTGCATGCCATCCCAACAAT---GCAGAGAAGAACTGCCACCTGCTTAAGA 3048  
3340 TGTGTTTTCCTTGCATCAAGCAGTTTCCGCAAAAGTCACTGATCCCATCAGACAA 3399  
3049 CCACTGACTGACTCAGGACTGTGTGTAAGACTCAGAGTGCAGCTGCTGCTCTGAT 3108  
3400 CCAGCTGATGAAGGAAGATCTTCCCTTAATGCTGTTGAGAAAGAACCTGTAACTGAAC 3459  
3109 GAAATTTGCTTCCAGTTCCAGTACAATCTAAACAGAGTGCAGACATGCCAGCTGAGTGA 3168  
3460 AAAAACTGCTACCACTATGTTATATATGTAACAGACTGAGTACATGTGCTGAGTGAG 3519  
3169 AACGCACTGTGGTCAAGGCTCAGGACCCGCTGCTAAGCTGTGTGTGCTGAGTATGCC 3228  
3520 AAGGCACTTTGTGGAATGAAATAAAAAACAAGATGTTGGAATGTTGTTTCAAGATGATGC 3579  
3229 AAGCCAGTCAGCATGGCAATGTGAGCAGCAATTTTGGAGAGGCCCCAGAGATGAGC 3288

Db 3580 AAGTCAGTTGACCTGAAATATTTGTGAAGCGCTTGGCTTGGAGAAACTGGCAGATGAAC 3639  
QY 3289 ATTCCCTGCTTGTGGAATGCGTGTCAACTGTCACTGTCACTGTCAAGGTGGAGCGCTTGACA 3348  
Db 3640 ACGTCTGCAATGTGGAATGCGCTGTGAACTGTCACTGTCACTGTCTGATTTGGTCTCCTTGGTCA 3599  
QY 3349 GAGTGTTCACAGACCTGTGGCCATGGAGTTCGAATGAGCGGACTCGATTTATCATTTATG 3408  
Db 3700 GAATGTTCTCAACATGTGGCTTCACAGGAAATATGATCCGAAGACGAACAGTGACCCAG 3759  
QY 3409 CCACCCAGGAGAGGAGCGCCATGCCCCACAGAGCTTACCAGGAGAAACCTGCCCA 3468  
Db 3760 CCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGACAGTCCAAACCTGCCCA 3819  
QY 3469 GTGACCCCTGCTACAGCTGGGTCCCTTGGCACTGTGTCTGCAATGTAATTTGGAGGTTGA 3528  
Db 3820 GTGAAGCCTTGTATCGGTGGCAATATGGCAGTGTCTCCATGCCAAGTGACGAGGCC 3879  
QY 3529 GACTGTGGGAAGAGTTCAGATCCGAGCCTTCTGATGCTGATGCTCCACAGTGGTTCAATA 3588  
Db 3880 CAGTGTGAGAGAGGACCCAGAACAAAGAACATTTCTTGTGTAGTAAGTATGAGGTGAGCT 3939  
QY 3589 TCTCATGCACTGGAACGTGTCGAGGATGCACTGTGTGGAGAAATGCCCTT----- 3638  
Db 3940 GATGATTTCAAGAAAGTGTGATGAGAAATCTGTGCTGACATTTGAATCATATAGAT 3999  
QY 3639 -----TCAGGACAGATCTGGAAGCAGCTGTGTCTGTGCTTGGCTTGGCCAGGAGCTGCCAT 3693  
Db 4000 GGTAAATAAAATATGTTCTGAGGAAATCTGAGCAGCGCTTGGCCAGGTGACTGTAT 4059  
QY 3694 TTAACAGAAATGTCAGAGTGGAGCAGATGTAATTAACCTGCAATGATGAAGAAGCTTT 3753  
Db 4060 TTGAAGACTGTTCTTCTGAGCCTGTGTGCTGACCTGACCTGTGATGATGAGTACTTA 4119  
QY 3754 GAGACTGTGGCCCGCAGCTAGATCAAGCACTTTTATTAATTCAGTCTTTTGTGAGAACAA 3813  
Db 4120 GCGTTTGGTGGAAATACAGTCCAGACCGGTGATTATACAGAACTAGAGAAATCAG 4179  
QY 3814 GACAGCTGCCCCCAACAGTCTTAGAAGACAGCGCTGTACAGAGGCAATGTTATCAC 3873  
Db 4180 CATGTGCCCCAGAGCAGATGTAGAAACAAATCATGTTATGATGACAGTGTCTATGAA 4239  
QY 3874 TACATGGAAGCAAGTCTTTGGAAACAATAAGCAAGAACTGTATGTCGCCAGCTTCA 3933  
Db 4240 TATAAATGGATGGCCAGTGTCTTGGAGGGCTCTTCCGAAACAGTGTGTTCAAGGTCA 4299  
QY 3934 GATGCGCTTATGTACAGAGGCTGCTCCCTCAGCGCCGTCTGCTGCTGCTGCTGCTGCTG 3993  
Db 4300 GATGTTATAATGTAAACAGGGGCTGCTTGGTATGAGCCAGCTGATGCCAGCTCT 4359  
QY 3994 TGCAATTCAGCTGCAGAAACCTTTCTCTCTGTACACAGGGTGGAGTCTGTGTTGT 4053  
Db 4360 TGTAACCCACCGTGTAGTCAACCCCTCTGTAATGTAGCGAGACAAACATGCCATTGT 4419  
QY 4054 GAGAAGGCTTATCAGAGATTAATAAATCAAAATGAGTGTGTTCTGGAATCTGCTGATGAAGTA 4113  
Db 4420 GAAGAGGCTACACTGAAGTCACTGCTTCTTAACAGCAGCCCTTGAGCAATGCACACTATC 4479  
QY 4114 CCAGGCTCAGAGGATATAAAGCTGATGTAATAAATCCTTTCTCGGAAACACAGAC 4168  
Db 4480 CCGTGTGTTGTTATPACCCCATGGAGACAAAGAGGAGATGTGAACCATGTC 4534

## RESULT 14

ACC72166  
ID ACC72166 standard; DNA; 4679 BP.

XX ACC72166;

AC 07-JUL-2003 (first entry)

XX Human NOV47e coding sequence.

XX

KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytosolic; nontropic; neuroprotective;  
KW antiparkinsonian; antileukemic; gene therapy; metabolic disorder;  
KW diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW immune disorder; haematopoietic disorder; Gens; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO2003029423-A2.  
XX  
XX 10-APR-2003.  
XX  
XX 02-OCT-2002; 2002WO-US031358.  
XX  
XX 02-OCT-2001; 2001US-0326483P.  
XX  
XX 05-OCT-2001; 2001US-0327342P.  
XX  
XX 09-OCT-2001; 2001US-0327917P.  
XX  
XX 09-OCT-2001; 2001US-0328029P.  
XX  
XX 09-OCT-2001; 2001US-0328044P.  
XX  
XX 09-OCT-2001; 2001US-0328056P.  
XX  
XX 12-OCT-2001; 2001US-0328849P.  
XX  
XX 15-OCT-2001; 2001US-0329414P.  
XX  
XX 17-OCT-2001; 2001US-0330142P.  
XX  
XX 22-OCT-2001; 2001US-0341058P.  
XX  
XX 24-OCT-2001; 2001US-0339286P.  
XX  
XX 24-OCT-2001; 2001US-0343629P.  
XX  
XX 29-OCT-2001; 2001US-0349575P.  
XX  
XX 01-NOV-2001; 2001US-0346357P.  
XX  
XX 12-APR-2002; 2002US-0371972P.  
XX  
XX 12-APR-2002; 2002US-0371980P.  
XX  
XX 17-APR-2002; 2002US-0373261P.  
XX  
XX 19-APR-2002; 2002US-0373805P.  
XX  
XX 23-APR-2002; 2002US-0374738P.  
XX  
XX 16-MAY-2002; 2002US-0381101P.  
XX  
XX 17-MAY-2002; 2002US-0381635P.  
XX  
XX 29-MAY-2002; 2002US-0383830P.  
XX  
XX 01-JUN-2002; 2002US-00362839.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
XX Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
XX Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
XX Rothenberg ME, Shinkets RA, Smithson G, Spytek KA, Taupier RJ;  
XX Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
XX  
XX WPI; 2003-381625/36.  
XX  
XX P-PSDB; ABR58454.  
XX  
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
XX treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
XX dyslipidemia, and in chromosome mapping, tissue typing or  
XX pharmacogenomics.  
XX  
XX Claim 20; Page 257-258; 487pp; English.  
XX  
XX The present invention relates to novel human NOV proteins and their  
XX coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
XX proteins are useful in manufacturing a medicament for treating a syndrome  
XX associated with a human disease. The NOV proteins and coding sequences  
XX may be used to diagnose, treat or prevent metabolic disorders such as  
XX diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
XX disorders such as Alzheimer's disease or Parkinson's disease, immune  
XX disorders, haematopoietic disorders and various dyslipidaemias  
XX  
XX Sequence 4679 BP; 1331 A; 1051 C; 1268 G; 1029 T; 0 U; 0 Other;  
XX  
XX Query Match 25.8%; Score 1133.4; DB 7; Length 4679;  
XX Best Local Similarity 57.7%; Pred. No. 0;  
XX Matches 2332; Conservative 0; Mismatches 1606; Indels 105; Gaps 13;  
XX  
XX 10 TGCATTGAGAGCTGAACCGAAGTGTGGTGCAGATGAAATGGAACACTTTGCCCTT 69

Db 288 TGATCCAGAAAGACAAAGACATTCCTGGGAGGATATCATCTGTGAGTACTTTGAGCCC 347  
Qy 70 CAGCCTCTCAGAAACAGGCTTGCTCATCTCTGTCCTCCCGGATTTGTGTAGTATCTGAG 129  
Db 348 AAGCTCTCTCCGGAGCAGGCTTGCTCATCTCTGTCCTCCAGCAAGATTTGATCTGTCTGAA 407  
Qy 130 TTCTTACCATGCTCCAACTGTAGCAAGGATGTGGAGAAATTCAGCATAGAACTCGC 189  
Db 408 TTTTCTGCTGCTCGAATGCTCCAGACCTGCGCAGCGGGCTCCAGCACCGAGCGGT 467  
Qy 190 GCGGTCAATAGCTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGAGTCAAGA 249  
Db 468 CATGTGGTGGCGCGCGCGAGTTCGGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCCAG 527  
Qy 250 GCTGTGATGCTCCCATTTCTCTGCTCTTGGGAAGAGGAATATACATTTAGCCTTTAAG 309  
Db 528 GTGTG-----CCAATCCAGTCCATGCGAGCGCGAGGCTCAGGTACAGCTGCAT 578  
Qy 310 GTTGACCAATGGAGTAAATGAGACTGCTCATTTAAAGAAATTAATCCAAAGCGGAGA 369  
Db 579 GTGGGGCTCTGGAGCACCTGCTCAATGCCCTCCACTCCGACAAAGTAAGACAAGCAGGAGA 638  
Qy 370 ACTGTTCTGGATTTTAA-----CTCTGATTCAATGAG 402  
Db 639 CCGGGAGAGATTAAGAACGGGAAGAGACCGCAGCAAGAGTAAGGATCCAGAGCC 698  
Qy 403 CGAGTCACTTTAAACATCAAAAGTTTACA-----AAGCACATCATTCATTCGAAGTCT 453  
Db 699 CGCGAGCTTTATTAGAAAAGAGAAACAGAAACAGGCAAGACAGACAAGAGACAATAAT 758  
Qy 454 TGGCAATAGATAGGTATCAACACCGGAGGTTTCTGTACAAAGATGATGACAA 513  
Db 759 TGGGACATCCAGATTTGATATCAGACACAGAGGTTATGTGATTAACAGACGGGAAA 818  
Qy 514 AATGCTATGTTAAGCCTTTGCTTCAAGATTCCTTCCCATTTGCTGCTGCTGCTGCTGCT 573  
Db 819 GCTGCTGATTAAAGCTTTTGGCAGCAAGAGAGCTTCCAATGACCTTCCAGTCTGTGTG 878  
Qy 574 ATGCCCAAGACTGTGAACCTCCAGTGGTCTCTCTGGAGCCCTGCTCCAGACATGC 633  
Db 879 ATACCAAAAGATGCGAGGTTTCCAGAGTGTTCAGAGTGGAGCCCTGCTCAAAAACATGC 938  
Qy 634 CGTTCAGGAGTCTCTTGGCAGGATTTAGGAGCAGGAGCGGAAACGTTGAAGCAGTGGCT 693  
Db 939 CATGACATGCTGCTCCCTCGAGGCTCTGTGTAAGACACGACCAACCATCAGGAGTTCC 998  
Qy 694 ATGGAGGTGGAAGAGAGTGTCTGAACTTCTTGAAAGAGAGGCTGCTGCTGTTGAAGA 753  
Db 999 ATTGGCAGTGAAGAGAGTGTCTCAGAAATTTGAAGAAAAGAACCTGTTGTTCTCAAGA 1058  
Qy 754 GAACTTCTGAGCAATGTCCAGGATTTCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 1059 GATGGAGTGTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1118  
Qy 814 GTCTCTCTCTCTGAGCAGGAGTATCCCACTGGCATGTGACGGGACCCCTGTTGCTGTC 873  
Db 1119 GTGGACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1178  
Qy 874 GTGGGATCCAGACCGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 921  
Db 1179 GGGGGATCCAGACCGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1238  
Qy 922 GCTGCGGCTGCTGAGGCGCAAGGAGTCTCTAGACCTGTGGAAAAGGCAATATGTGTGGA 981  
Db 1239 TTAAGTACCACCAAGAAACAAAGAGGCTCAAAGCCAAATGGAATTAATATGCTGCTGGA 1298  
Qy 982 CCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1041  
Db 1299 CCTATCCCTAATATACAGAGTGTGCCCATTCCTTGTCCCACTGAATGTGAAGTTTCA 1358  
Qy 1042 TCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1101

b	1359	CCTTGGTCAGCTTGGGGACCTTGTACTTATGAAAACTGTAATGATCAGCAAGGGA AAAA	1418
y	1102	GGATTTAGAACGAGGCAGGG-----CCATGTCTCATGGAATCTACAGGGCTGCA	1152
b	1419	GGCTTCAAACGTAGGNAAGGGCGCATTACCAATGAGCCCACTGGAGGCTCTGGGTAACC	1478
iy	1153	GGGCATTGCCCTCATTTGGTGGAGTCTGTTCTTGTGAGGATCCAAATGTGCTACCGATGG	1212
ib	1479	GGAAACTGCGCCCTCACTTACTGGAAGCCATTCCCTGTGAAGAGCCTCCCTGTTATGACTGG	1538
iy	1213	---CTGGCATCAGAAGGGATCTGTTTCCCTGATCATGGAA---AATGTGGCTGGGACAT	1266
ib	1539	RAAGCGGTGAGACTGGGAGACTGCGAGCCAGATAACGGAAAGGAGTGTGGTCCAGCGACG	1598
iy	1267	CGTATTTGAAGCCGTCTGCGAGAAATGACCGGGAGAAAGATGTATCAGGGAGTCTTTTGC	1326
ib	1599	CAAGTTCAAGAGGTTGTGTGCATCAACAGTGTATGGAGAAGAAGTTTACACAGACGTGTGC	1658
iy	1327	CCAGTTCCCCCTCTCTCTGAGAGGAAGTCTGTGAAATTTCCCTGCGCAATGGACTCTGTG	1386
ib	1659	AGAGATGGCATCTTCTCCCATCCCTGTGGCTGTGATGCCCCATGCCCGAAAGACTGTGTG	1718
iy	1387	CTGAGCGAGTGGACGGAGTGGTCATCTCTGTTCCTAGTCTCTGTTCAAATAAAAACTCAGAT	1446
ib	1719	CTCAGCACATGTTCTACGTGGTCTCTCTGCTCACACACCTGCTCAGGGAAAAACGACAGAA	1778
iy	1447	GGGAAACAGACACAGGTCAAGAACTATCTTGGC---ACTGGCTGGGAAAGGTGAAGAGCCA	1503
ib	1779	GGGAAACAGATACGAGACAGATCCATCTCTGGGCTATGCGGGTGAAGAGGTGGAATTCG	1838
iy	1504	TGTCCCCCTAGTCAGGCTCTCAAAGAGCATCGTTGTGTAATGACCATTCCTGTATGCGAG	1563
ib	1839	TGTCCAAATAGCAGTGCTTTGCAAGAAGTACGAAGCTGTAATGAGCATCTTGTCACAGTG	1898
iy	1564	CTTCACTGGGAGACATCGCCTTGGGGCCCTTGTCTTGAGACACATTTGTPACTGCGCTT	1623
ib	1899	TACCACATGGCAAACTGGTCCCTGGGGCCAGTGCAATTGAGACACCTCAGTATCGTCTTTC	1958
iy	1624	AATGCAACCATTTGGCTGGAATGAGAAAGCCACGTGGTGTAGGCATTCAAGCTCCGAGA	1683
ib	1959	AACAACAATTACGACTTGGAAATGGGAGGCTCTCTGCTGTGCGCATGCAGACAAGAAA	2018
iy	1684	GTCTTCTGTCTCAAGAGTCAGTGGGACAAAGTAATGACCAAAGATGTCCAGATTTACT	1743
ib	2019	GTCACTGTGTGGAGTCAATGTGGGCCAAGTGGGACCCCAAAAAATGTCTGAAAGCCTT	2078
iy	1744	CGACCTGAAACTGTGCGCCCTCTTTTCTCCCATGCAAAAAAGACTGTATTGTGACTGCT	1803
ib	2079	CGACCTGAAACTGTAGGCCCTTGTCTGCTTCTTGTGTGTAAGAGACTGTATTGTGACCCCA	2138
iy	1804	TTCAAGTGAAGTGGACACCCTGCCCAAGGATGTGCCAAGCAGGAAATGCCAGT---AAAA	1860
ib	2139	TATAGTACTGGACATCATGCCCTCTTCTGTGTAAGAAGGGACTCCAGTATCAGGAAG	2198
iy	1861	CAGTCTCGATACAGATCATCATCAAGANGCAGCCAAATGGAGCCAGGAATGCCCAGAT	1920
ib	2199	CAGTCTAGGCATCGGGTTCATCATCTCAGCTGCCAGCAACCGGGGGCCGAGACTGCACAGAT	2258
iy	1921	ACCTTATATCAGGAGAGAGAGTGTGAAGATGTTTCCCTGTGTCTGTATATCCGTGGAAG	1980
ib	2259	CCCCCTCTATGAAGGAGGCTGTGAGGCACCTCAAGCGTGCCAAAGCTACAGETGGAAG	2318
iy	1981	CCAAGAAATGGAGCCCTTGCACTTTAGTGCCAGAGTCTGTCTGGGAGGGAATAACGGGC	2040
ib	2319	ACTCAAAATGGGCGAGATGCCAATTAGTCCCTTGGAGCGGTGCAACAGACAGCCCTTGGGA	2378
iy	2041	AGCAGTGAAGCTGTGTGAAAGGGGTTTACAAACAAAGAGCTGTCTCATCATCTCTGTATGAC	2100
ib	2379	GCACAGGAAGGCTGTGGGCTGTGGGCGACAGGCAAGAGCCATTACTTGTCCGACAGCAT	2438
iy	2101	AACCGGTACAGAAATGATGGAATGCCTCAAGCAGACAAACGGCAATGCTCTCTCTTGTG	2160
ib	2439	GGAGGACAGGCTGGAATCCATGAGTGCGCTACAGTATGTCAGGCCCTGTGCGAGCCCTTACC	2498

2161	QY	CAAGAAATGCACAGTCCCACTGTGCGAAGAGCTGCACCTTCACTGCTTGGTCCAGTTTACG	2222
2499	DB	CAGGCTGCCAGATCCCTGTCAGAGTAGCTGCAATTGACACGCTGGTCCAAAGTTTCT	2558
2221	QY	CCCTGTGCCACGAACGTGAAGCCACAAAAGTAGCGCGCAGCAGCTCACAGGAAAAGC	2280
2559	DB	TCATGCAATGAGACTGTGTGCAGTTAGGACACGAAAGCGCACTCTTGTGTGAAAAGT	2618
2281	QY	AGAAAGAGAGAAATGCCAGGATTCGACCTTTACCTCTAGTGGAGACAGAACTATGT	2340
2619	DB	AAAAGAGGAAAAATGTAAAAATCCCATTTGTATCCCTGATTCAGACTCAGTATTGT	2678
2341	QY	CCTTGTGATGAAATTTATATCCCAACCTTATGGAACCTGTCAGATTGCATTTCTTCAGAA	2400
2679	DB	CCTTGTGACAAATATATGACACAACCTGTGGGAACTGGTCAGACTGTATTTTACCAGAG	2738
2401	QY	GGCAGAAGGAGGCTCACCGAGACTGCGGGTACAGCGACAGCAAGAAATCTGGAGAA	2460
2739	DB	GGAAAAGTGAAAGTGTGTGGGAAATGAAGTAGCAAGGAGACATCAAGGAATCGGACAA	2798
2461	QY	GGCTCGGCTTTCCGAGCAGTAGGCTGTTCTGATAAAAAATGGAAGACCTGTTGACCCCTCC	2520
2799	DB	GGATATCGTTACCAAGCAATGGCATGTCAGATCAAAATGCGAGGCTGTGTGAAAACATCT	2858
2521	QY	TTCTGCAGCAGCTCTGGTTACATTCAGAAAATATGTCAATTCCTCGCCCATTTGATTCG	2580
2859	DB	AGATGTAAACAGCCATGGTTACATGTAGAGGCGCTGCATCATCCCTGCCCCCTCAGACTGC	2918
2581	QY	AAGTTAAGCAATTTGGTCTAGTTTGGGGTCTTTCGAGTTTCATTTTGTGGAATGGAGTGAGA	2640
2919	DB	AAGCTCAGTCAGTGGTCCAACTGGTCGCGCTGCAGCAAGTCTCTGTGGAGTGGTGTGAAG	2978
2641	QY	ATTGATCCCAATGGCTTAAAGAAAACCTTCACATGGAGGACGACCATCTCCCAACTG	2700
2979	DB	GTTCTGTTCTAAATGGCTGCGTGAAGAACCATATATATGGAGAAAGCCCTTGCCCCAACTG	3038
2701	QY	GATC-----TCAAGAAATCAGGTACATGAGGCACTGCCATGTTTACAGTGAAGTGCATCAG	2754
3039	DB	GACCATGTCAACAGGCACAGGTGTATGAGTTGTCCATGCCACAGTGACTGCAACAG	3098
2755	QY	TATTCCTGGTGTAGAACACTGGTCTTCATGCAAAATCAACATGAGTGAAGTGCCTG	2814
3099	DB	TACCTATGGGTACAGAGCCCTGAGGCATCTGCAAGTGACCTTTGTGAATATGCGGGAG	3158
2815	QY	CGCTGTGGAGGAGAAACCAATCTAGAAAATCAGATGTGT---GAATATCGGGATGGT	2871
3159	DB	AACCTGTGGAGGGCGTGCAAAACCGAAAAGTGAGATGCATGCAGAATACAGCAGATGGC	3218
2872	QY	GAAAGTGGAGCAGTGATAGCAACTGTGCAACCCAGGATGAAATCCCCACAGAAACCCAG	2931
3219	DB	CCTTCTGAAATATGATAGAGATTACCTCTGTGACCGAAGAGATGCCCTTGGGCTCTAGA	3278
2932	QY	TCCTGTTCTCTATGTGCCAATCTAGTGTGTCAATGTCTGAGTGGGACATTTGGAGCAAA	2991
3279	DB	GTGTGCAAAATACCATGCCCTGAGGACTGTGTGATATCTGAATGGGGTCCATGACCCCAA	3338
2992	QY	TGCCCACTGATCCGATCCCAACAAAT--GCAGAGAAGAACTCCGCCACTGCTTAAGA	3048
3339	DB	TGTGTTTGGCTTGCATCAAAGCAGTTTCGGGCAAGAGTCAAGTATCCCATCAGACAA	3398
3049	QY	CCATCACTGAATCAAGGACTTGTGTGAAGACTCAACAGTGCAGCTTGGCTCCTGAAT	3108
3399	DB	CCAGCTGATGAAGGAAGATCTTGCCCTTAATGCTTTGAGAAAGAACCCCTGTAACTGAAC	3458
3109	QY	GAAATATGCTTCCAGTTCCAGTACAATCAACAGAGTGGAGCACATGCCAGCTCAGTGAA	3168
3459	DB	AAAACTGCTACCACTATGATTATATGTAACAGCTGGAGTACATGTCACTGAGTGAG	3518
3169	QY	AACGCACCGCTGTGTCAAGGCGTCAGGACCGCCCTGTCTAAAGTGTGTGTGTCAGTGTGCG	3228
3519	DB	AAGGCAGTTTGTGAAATGGAATAAAAACCAAGATGTTGGATTGTGTTCGAAGTATGGC	3578





1267 CGATTCTGAAGCCCTCTGCCAGAAATGACCGGAGAAAGATGTATCAGGGAGTCTTTGC 1326  
2041 CAAGTTCAAGAGGTTGTGTGATCAACAGTGATGGAGAAAGTTGCACAGACAGCTGTGC 2100  
1327 CCAGTTCCCTCTCTCTGAGAGAGTCTTGTGAAATTCCTTCCGCAATGAGTCTGTG 1386  
2101 AGAGATGCCATCTTCCCATCTCCCTGTGGCTGTGATGSCCCCATGSCCGGAAAGACTGTG 2160  
1387 CTGAGCAGTGGAGCGAGTGGTCTATCTCTGTTCCTCAGTCTCTTCAAAATAAAACTCAGAT 1446  
2161 CTCAGCACATGGTCTAGTGGTCTCTCTCTGCTACACACCTGCTCAGGAAACAGACAGAA 2220  
1447 GGGAAACAGACAGAGTCAAGAACTATCTGGGCACTGGCTGGG--GAAGTGAAGACCA 1503  
2221 GGGAAACAGATACAGACAGATCCATCTTGGCCCTATGCGGGTGAAGAGGTGAATTCG 2280  
1504 TGTCCCTCTAGTCCAGGCTCTCAAGAGCATCTTGTGTAAATGACCATCTCTGTATGCAG 1563  
2281 TGTCCAAATAGCAGTCTTGTCAAGAGTACGAAGCTGTAAATGAGCATCTTGCACAGTG 2340  
1564 CTTCATCTGGAGACATCGCTTGGGCTCTTGTCTTCTGAGGACATCTGGTAACTGCGCTT 1623  
2341 TACCACTGGCAAACTGCTCTCTGGGCTGAGTCAATTCAGGACACCTCAGTATCGTCTTC 2400  
1624 AATGCAACCATTTGCTGGAATGGAGAGCCAGCTGTGTGTAGGCACTTCAGACTCCGAGA 1683  
2401 AACAACTAGCACTTGGAAATGGGAGGCTCTCTGCTGTCTGTGCGCATGCGACAGAAA 2460  
1684 GTCTTCTGTCTCAAGATCAGCTGGGACAGTAAATGACCAAAAGATGTCCAGATCTTACT 1743  
2461 GTCATCTGTGGAGTCAATGTGGGCCAAGTGGGACCCCAAAATGTCTGAAAGCTT 2520  
1744 CGACCTGAAGTCTGGCCCTCTTCTCCCATGCAAAAGACTGTATTGTGACTGCT 1803  
2521 CGACCTGAAGTCTGAAAGCTTGTCTCTCTTCTTGAAGAGACTGTATTGTGACCCCA 2580  
1804 TTCACTGAGTGGACACCTCTCCCAAGATGTGCAAGAGGAAATGCCACAGT---AAAA 1860  
2581 TATAGTACTGGACATCATGCCCTCTCTGTGTAAAGAGGGGACTCCAGTATCAGGAAG 2640  
1861 CAGTCTCGATACAGATCATCTCAAGAGAGCCAAATGGAGGCCAGGAATGCCAGAT 1920  
2641 CAGTCTAGGCATCGGGTCACTTCACTGCTGCCCAACCGGGGCGGAGACTGCACAGAT 2700  
1921 ACCTTATCAGCAGAGAGAGTGTGAAGATGTTTCTGTGTCTCTATATCGGTGGAAG 1980  
2701 CCCTCTATGAAGAGAGGCTGTGAGGACCTCAAGCGTGCACAAAGCTACAGTGGGAAG 2760  
1981 CCACAGAAATGGAGCCCTTGCACTCTTGTGCGCAGAGTCTGTCTGGCAGGGAATAACGGGC 2040  
2761 ACTACAAATGGGCGAGATGCCAATAGTCTCTTGGAGGCTGCAACAGACAGCCCTGGA 2820  
2041 AGCAGTGAAGCCCTGTGGAAGGGGTTACAAACAGAGCTGTCTCATGTCATCTCTGATGAC 2100  
2821 GCACAGAGGCTGTGGGCTGGGCGACAGGCAAGAGCCATTACTTGTGCAAGAGAGAT 2880  
2101 AACCGGTTCAGCAGAAATGATGGAATGCCCTCAAGCAGACAAACGGCATGCTCTCTTGTG 2160  
2881 GGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGAGGCCCTGTGCGAGCCCTTACC 2940  
2161 CAAGATGCAAGTCCCATGTCGAGAGAGTGCACCTTCACTGCTTGGTCCAGATTTAG 2220  
2941 CAGGCTGCGAGATCCCTGCGCAGGATGATGTCAATGACCAAGCTGGTCCAAAGTTTCT 3000  
2221 CCCTGCTCCCAACTGTGAAGCCCAAAAGTAGGCGCG 2261  
3001 TCATGCAATGGAGCTGTGGTGTAGGACCCAGAAAGCG 3041

Search completed: February 23, 2004, 12:24:16  
Job time : 1583.36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 23, 2004, 11:46:56 ; Search time 10047.4 seconds  
(without alignments)  
13071.481 Million cell updates/sec

Title: US-10-022-710-1

Perfect score: 4398

Sequence: 1 atggcgctgcattcagaa.....atggagacttagacatgtaa 4398

Scoring table:

IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_man:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_pbg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1963.4	44.6	3897	11	AK051714 Mus muscu
2	1908.8	43.4	3242	11	AK083798 Mus muscu
3	866.4	15.2	2266	11	AK047508 Mus muscu
4	621.8	14.1	676	10	BE299433 Mus muscu

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

5	600.4	13.7	760	10	BE295719
6	565.2	12.9	772	14	CF729699
7	543.4	12.4	757	14	CB723142
8	542.6	12.3	820	13	BU703731
9	506.8	11.5	717	13	BY735628
10	485.8	11.0	683	10	BB451562
11	472.6	10.7	600	12	BI988413
12	451.6	10.3	565	9	AV590588
13	447.2	10.2	713	12	BM414542
14	446	10.1	551	10	BE899999
15	422.4	9.6	600	12	BI981468
16	421.6	9.6	600	13	BU251239
17	404.8	9.2	491	10	BF605229
18	354.2	8.1	670	10	BB616046
19	330.6	7.1	693	13	BU313848
20	288.2	6.6	531	9	AL713171
21	283.4	6.4	676	14	CF531936
22	272.6	6.2	535	10	AW787130
23	243.2	5.5	760	13	BU262334
24	241	5.5	735	13	BU283850
25	239.4	5.4	750	13	BU308368
26	239.4	5.4	793	13	BU200855
27	223	5.1	658	13	BU267027
28	224	4.9	944	13	BU366211
29	212.6	4.8	316	10	BA452816
30	201	4.6	828	28	AQ531624
31	197	4.5	738	13	BU451181
32	196.2	4.5	556	10	BE232461
33	195.2	4.4	649	14	CF362876
34	194.8	4.4	801	13	BU318139
35	194.6	4.4	523	12	BI340182
36	189.4	4.3	503	10	BF441768
37	189.2	4.3	673	28	AQ282877
38	184.8	4.2	531	28	AQ737047
39	181.4	4.1	677	10	BB463698
40	179.4	4.1	529	12	BI340184
41	178.2	4.1	928	13	BU267473
42	168	3.8	695	13	BU451889
43	167.6	3.8	583	13	BX100582
44	166	3.8	438	10	BF441422
45	165.6	3.8	475	10	BF441417

## ALIGNMENTS

RESULT 1  
AK051714  
LOCUS  
DEFINITION  
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length enriched library, clone: D130067103 product: weakly similar to HYPOTHETICAL PROTEIN DJ1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens], full insert sequence.

ACCESSION  
AK051714

VERSION  
AK051714.1

KEYWORDS  
HTC; CAP trapper.

SOURCE  
Mus musculus (house mouse)

ORGANISM  
Mus musculus

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS  
Carninci, P. and Hayashizaki, Y.

JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)

MEDLINE  
99279253

PUBMED  
10349636

REFERENCE  
1

AUTHORS  
Carninci, P. and Hayashizaki, Y.

JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)

MEDLINE  
99279253

PUBMED  
10349636

REFERENCE  
2

AUTHORS  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)

MEDLINE  
99279253

PUBMED  
10349636

REFERENCE  
3

AUTHORS  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)

MEDLINE  
99279253

PUBMED  
10349636

11042159

PUBMED  
REFERENCE  
AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hahama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--394-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

20530913  
11076861  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

JOURNAL  
REFERENCE  
AUTHORS

5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

JOURNAL  
REFERENCE  
AUTHORS

6 (bases 1 to 3897)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hasegaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiracka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission

TITLE  
JOURNAL

Submitted (18-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp].  
URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>.

FEATURES  
source

Location/Qualifiers  
1..3897  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM.DB:D130067103"  
/db\_xref="MGI:2420048"  
/db\_xref="caxon:10090"  
/clone="D130067103"  
/tissue\_type="spinal ganglion"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="12 days embryo"  
174..3497  
/note="unnamed protein product; putative  
weakly similar to HYPOTHETICAL PROTEIN D0110N13.1 IN  
CHROMOSOME 7 (FRAGMENT) [Homo sapiens] (SPTR|O43376,  
evidence: FASTA, 53.6%ID, 99.4%length, match=540)"  
/codon\_start=1  
/protein\_id="BAC34732.1"

## CDS

```

/db_xref="GI:26342140"
/translation="MFLRSDLAVTHWYSRNRKFLVLSLLSOAAHLEGRKDNQFLW
KTGWGACGDCGPGQRAVWCFHLEGWTSMSNDESSQPPKERSCFRVDMDHSD
LFQWEVSDHRCILLPGAGQEPRAVECVTAQGLQHRTVRCLOKLNRWTSNEICE
HPAPOPTEOACILPCPRDQVSEFSPWSTCPGCGKLOHRTVEVTAAPLYGSLQCE
NLTSERACEAPVSPKESYSLKVPKSKLPHLKEVDLSGRNIQDFSSDSNQ
VLTHTSKAHHSPQGVVIGFQTRQVWCTRSRGNALLSLCVRDSEFFLTVQPCVMP
KXDTSSWSPSCPKTSGTLLSPGVRSSRNKVIHAGGQPCPEVREYCAQSLPATI
EFLQPCRYSPVRLTCLGPAPASQOLCNVPCMDCVISWSWTGFCVFNCHDPOGKK
ASRTKEVSRPVRTLCGLPAPASQOLCNVPCMDCVISWSWTGFCVFNCHDPOGKK
KPRVQRHLMESTGPMGRCPHLAESVPCDPMCHRWLASEGICIAHDGKGLGHRIL
KAVCONSERGVSGGLQCPVPPPERMAEIPCRMDCVVSMTVWSSCSQSCSNKNSDG
KQTRSRILALAGCGKTCPSSELOEYRLNDHSCOTLYWETSAMGSCSENTLVTL
NVTIGWGEATCGVGIQTRKVFQIKSHVQVMKRCPESTRPETHVRCFLPKCKDCLV
TAFSEWTPCPRCPQGNITIKQSYRIIIQEAANGQCEPDILFEERCEDEKSLQNG
RWKQWSSCILVPEDSIRQRTSEACGLQTRAVSCISDDNQSAEMTECLQMGNG
MPVLVQECTIPCRDCTFTWPKFTPCSKCEATQIRRLQTKGRSKCKQDASLYP
LVMEPEPCDFTMSHPYNSACILPEGRKDAQOGLVQGDCKECCGEGVRFRAIAGS
NINRGPRVDPSPCNSSGVQIEACVIPCPCDKLSDWSWSSGSCSSCGIGVIRSKWLKE
KPYSGRPRCPKLDLKNQFCHLSISSGATLFLKILLMDMSFLFAVVCHFCYFQFLF
ICVFQRLVHQVFRDLCA"
polya_signal
3880..3885
/note="putative"
polya_site
3897
/note="putative"
ORIGIN
Query Match      44.6%; Score 1963.4; DB 11; Length 3897;
Best Local Similarity 82.9%; Pred. No. 0;
Matches 2253; Conservative 0; Mismatches 461; Indels 3; Gaps 1;
QY      4 GTGCCTGCATTGCAAGCTCAACCGAACTGTGGTTGCAAAATGAAATGCGAACACTTT 63
DB      600 GTGCCTGTCTTCAGAACTCAATAGAACCATGGTGTCCATGAGATCTGTGAACACTTT 659
QY      64 GCGCTTCAGCTCTTACAGACAGAGCTTGCCTCATCTCTTGTCCCGGGATGTGTAGTA 123
DB      660 GCTCCTCAGCCCCCACCAGACAGAGCTGCTCATCTTCTCCCGCAGGACTGGTGTGTA 719
QY      124 TCTGAGTTCTTACCATGTGTCACACTGTAGCAAGGATGTGGAGAAATTCAGCATAGA 183
DB      720 TCTGAGTTCTCCCAATGTTCACCTGCCCCAGAGATGTGGAGAAATTCAGCATAGA 779
QY      184 ACTCGCGGGTCATAGCTCCCCCTCTCTTTGGTGGTTTGCATGTGCAAAATCTGACTGAG 243
DB      780 ACTCGTGTGGCCATCGCGCCCTCTGTATGAGAGTCTGCAGTGTCCAAATCTCACGGAG 839
QY      244 TCAAGAGCTGTGATGCTCCCATTTCCCTCTCTTGGGAGAGGATATACATTTAGC 303
DB      840 TCCAGAGCTCGGAGGCTCCAGTTTCCCTCTCTTGGGAAAGAGATATTCCTTCAGC 899
QY      304 CTTAAGGTTGGACCATGGAGTAAATGCAGACTGCCTCATCTTAAAGAAATTAATCCAGC 363
DB      900 CTTAAGGTTGGACCATGGAGCAATGTAGACTCCCTCACCTTAAGGAAGTCGACCTCAGC 959
QY      364 GGAAGACTGTTCTGGATTTTAACCTCTGATTCAAATGAGCGAGTCACCTTTAAACATCAA 423
DB      960 GGAAGAAATATCCAGGATTTTAGCTCTGATTCAAACGAGCAGTCAACCTAACCATCAG 1019
QY      424 AGTTTCAAGACCATCATCATTCGAAGTCTTGGGCAATAGAGATAGGTTATCAAAACCCGG 483
DB      1020 AGTTTCAAGACCATCAACCATCCCGCTGGGATGTAGTAGAGGTTTCAAAACCCAGG 1079
QY      484 CAGGTTTCGTGTACAGAGATGATGCAAAATGCTATGTTAAGCCTTGGCTTCAAGAT 543
DB      1080 CAGGTTGTGTGTACCAAGAGTGTATGGAAGAAACGCGCTTTTAAAGCCTTTGTGTGAT 1139
QY      544 TCCTTCCCATTCAGCTGTTCCAGTCTTCGATCCCAAGACTGTGAAACCTCCAGTGG 603
DB      1140 TCCTTCCCTCTGACTGTCCAGCTTGTGTATGCCAAAGACTGTGAGACTTCTGAATGG 1199
QY      604 TCCTCTCGAGGCCCTTGCTCCAGACATGCGGTTCAGGAGTCTCTTCCAGGATTTAGG 663

```

1200	TCTCCATGGAGCCCTGTTCCAGACCTGCGCTTCTGGACCTTTATCACTCGGAGTTAGA	1259	CGACCTGAAACTGTGCGCCCTGTTTCTCCCATGCAAAAAAGACTGTATTGTGACTGCT	1803
664	AGCAGGACCGGAACGTTGAAGACAATGGCTATTGGAGTGGAAAGAGTGTCTGAACTTT	723	CGGCCAGAACTGTGAAGACCTTGTGTTTCTCCATGCAAGAAAGACTGCTTGTGACTGCT	2399
1260	AGCAGGACCGGAACGTTGAAGACAATGGCTATTGGAGTGGCAAGATGCCAGAACTT	1319	TTCACTGAGTGGACACCTGCGCCACGCGCCCTCCAGAGCCAGGAAATACTACATAAACA	1863
724	CTTGAGAAAGAGGCTGCTGATTTGTAAGAGAACTTCTGCAAGCAATGTCCTGAGTATTC	783	TTCACTGAGTGGACACCTGCGCCACGCGCCCTCCAGAGCCAGGAAATACTACATAAACA	2459
1320	CTTGAGAAAGAGGCTGCTGATTTGTAAGAGAACTTCTGCAAGCAATGTCCTGAGTATTC	1379	TTCTCGATACAGAAATCATCATCAAGAAGCAGCAATGGAGGCCAGGAATGCCCAGATAC	1923
784	TGGAACTTCTGAATGGAAGAATGCCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	843	TTCTCGATACAGAAATCATCATCAAGAAGCAGCAATGGAGGCCAGGAATGCCCAGATAC	2519
1380	TGGAACTTCTGAATGGAAGAATGCCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1439	TTATATGAGAGAGAGTGTGAAGATGTTTCTTGTGTCTCTGATATATCGTGGAGGCA	1983
844	CACTTGGCATGTGACGGGACCGCTGTGCGGTGGATCCAGACCCGGGAGTGTACTGT	903	TTCTGATACAGGATTTATCATCCAGGAAGCAGCAATGGAGGCCAGGAATGCCCAGATAC	2579
1440	CTTTGGCATGTGACAGGACCAATCTGTGTGTGGTGGATCCAGACCCGGGAGTGTACTGT	1499	TTCTGATACAGGATTTATCATCATCAAGAAGCAGCAATGGAGGCCAGGAATGCCCAGATAC	2579
904	GCCAGAGGCTTACCAAGAGCTGCGCACTGAGGGCCAAAGGAAGTCTCTAGACCTGTGAA	963	CAGAAATGAGGCCCTTGTGATCTTAGTGCAGAGTCTGTCTGGCAGGGAATAACGGGACG	2043
1500	GCCAGAGGCTTACCAAGAGCTGCGCACTGAGGGCCAAAGGAAGTCTCTAGACCTGTGAA	1559	CAGAAATGAGGCCCTTGTGATCTTAGTGCAGAGTCTGTCTGGCAGGGAATAACGGGACG	2639
964	AGGCAATATGTGTGGGACCCCGGTTGCCCTCTCAGCTCTGCAATATCCCTTGCTCT	1023	AGTAAAGCCCTGTGGAAGGGGTTACAAACAAGAGCTGTCTCATGCACTCTCTGATGACAAC	2103
1560	AGGCAATATGTGTGGGACCCCGGTTGCCCTCTCAGCTCTGCAATATCCCTTGCTCT	1619	AGTGAGGCATGTGGGAAGGGGCTGCAGACAAGAGCTGTCTCATGCACTCTCTGATGACAAC	2699
1024	ACGACTGATATGTTCTCTGCTGAGCTGCGGCTGTGATCCATGCAAACTGTGAT	1083	CGCTCAGCAGAAATGATGGAATGCTTCAAGCAGACAACAGGCAATGCTCTCTCTGTGCAA	2163
1620	ATGGAATGATATGTTCTCTGCTGAGCTGCGGCTGTGATCCATGCAAACTGTGAT	1679	CAGTCTGCAAAATGATGAGTGTCTTCAAGCAGATGATGATGCAATGCTCTCTCTGTGCA	2759
1084	GAACCTCAGGGGAAAGAAATTTAGAACAGGAGCGGCTGTCTCATGCAATCTACA	1143	GAATGCAAGTCTCCCTGTGCGGACGACTGCACCTTCAACCTTGGTCTTAACTTCACTCC	2819
1680	GATCTCAGGGGAAAGAAATTTAGAACAGGAGCGGCTGTCTCATGCAATCTACA	1739	TGCTTCAAGAACTGTGAAGCAACTCAGATCAGAGGCGGCGCTTACAGGGAAGAGCAGG	2879
1144	GGGCTGCAAGGCAATGCTCTCTGCTGAGTCTGTTTCTCTGAGGATCCAACTGTG	1203	AAGAAGAGAAATCCAGGATTTGACCTTTACCTCTAGTGGAGACAGAACTATGTCCT	2343
1740	GGGCTGCAAGGCAATGCTCTCTGCTGAGTCTGTTTCTCTGAGGATCCAACTGTG	1799	AAGAAAGAGAACTGCCAGGATGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2939
1204	TACCGATGCTGCAATCAGAAAGGATGCTGTTTCTCTGATCATGGAATAATGCGCTGGA	1263	TGTGATGATTTTATCCCAACCTTATGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2403
1800	CATCGTGGCTTCCCTCGAAGGATCTGTTTCTGCTGATCATGGAATAATGCGCTGGA	1859	TGTGATGATTTTATCCCAACCTTATGGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2999
1264	CATCGTATCTGAGGCGCTGCGAGATGACCGGAGAGATGATCATGAGGATCTT	1323	AGAGGAGAGCTTCACTCCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2460
1860	CATCGATTTCTGAGGCTGCTGTTGATGATGAGGAGGAGGATCATGAGGATCTT	1919	AGAGGAGAGCTTCACTCCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	3059
1324	TGCCAGTCTCCCTCTCTGAGAGAGTCTGTTGAAATTCCTGCGCAATGGAATGCT	1383	AGAGGAGAGCTTCACTCCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	3119
1920	TGTCAGTCTCTCTCTGAGAGAGTCTGTTGAAATTCCTGCGCAATGGAATGCT	1979	AGAGGAGAGCTTCACTCCAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	3179
1384	GTGCTGAGGAGTGGAGGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1443	TTCTGACGAGCTGTGTTACATTTCAAGAAATGCTGCTTCTCTGCTGCTGCTGCTGCTGCT	2580
1980	GTAGTGTGAGTGGAGGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2039	TTCTGACGAGCTGTGTTACATTTCAAGAAATGCTGCTTCTCTGCTGCTGCTGCTGCTGCT	3179
1444	GATGGGAAACAGACAGCTCAAGAACTATCTCTGCACTGCTGCGGAGGAGTGGAAAGCA	1503	AGAGTAAAGCAATGCTGTTAGTGGGCTTGTGCACTTCACTTGTGGAATTTGGAGTGA	2640
2040	GATGGGAAACAGACAGCTCAAGAACTATCTCTGCACTGCTGCGGAGGAGTGGAAAGCA	2099	AGAGTAAAGCAATGCTGTTAGTGGGCTTGTGCACTTCACTTGTGGAATTTGGAGTGA	3239
1504	TGTCCCTCTAGTCAAGCTCTCCAGAGATGCTGTTGTTGATGACCAATCTCTGATGCA	1563	AGAGTAAAGCAATGCTGTTAGTGGGCTTGTGCACTTCACTTGTGGAATTTGGAGTGA	2700
2100	TGCCCTAGTGTAGGAACTCCAGAGTACCGCTATGCAATGATCATCTCTGTACACAG	2159	AGAGTAAAGCAATGCTGTTAGTGGGCTTGTGCACTTCACTTGTGGAATTTGGAGTGA	3299
1564	CTTCACTGGGAGATCGCTTGGGCGCTTGTCTGAGGACATTTGGTAACTGCGCCTT	1623	GATCTCAAGAACTCAGGT 2717	
2160	CTTCACTGGGAGATCGCTTGGGCGCTTGTCTGAGGACATTTGGTAACTGCGCCTT	2219	GATCTCAAGAACTCAGGT 3316	
1624	AATGCAACCTTGGCTGGAATGGAGAGCCAGTGTGTTGTTGAGGATTCAGACTGAGAG	1683		
2220	AATGTGACCAATGGCTGGAATGGAGAGCAACATGTGGGCTGGGCTTACAGACGAAAG	2279		
1684	GTCTTCTGTCTGAGAGTCAAGTGGGCAAGTATGACCAAAAGATGTCCAGATTTCTACT	1743		
2280	GTCTTCTGTCTGAGAGTCAAGTGGGCAAGTATGACCAAAAGATGTCCAGAGTCTTACA	2339		

RESULT 2  
AK083798  
LOCUS

AK083798 3242 bp mRNA linear HTC 20-SEP-2003



1246 AGCAGGAGCCGGAACGTGAAGCATATTGCTATTGGAGGTGGCCAGGAATGCCAGAACTT 1305  
2Y  
724 CTTGAGAAAGAGGCGCTCATTTGTTGAAGGAGAACTTCTGACAGAAATGTCAGGTAATCC 783  
Db  
1306 CTTGAGAAAGAGGCGCTCATTTGTTGAAGGAGAACTTCTGACAGAAATGTCAGGTAATCC 1365  
2Y  
784 TGGAGAACTTCTGAAATGGAAGAAATGCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 843  
Db  
1366 TGGAGAACTTCTGAAATGGAAGAAATGCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1425  
2Y  
844 CACTGGCATGTGACGGGACCGCTGTGTGGCGGTGGATCCAGACCCGGGAGGTGTACTGT 903  
Db  
1426 CTTTGGCATGTGACGAGACCACTGTGTGTGGGTGGATCCAGACCCGGGAGGTGTACTGT 1485  
2Y  
904 GCCCAGAGCGTACAGCAGTGTCCGCACTGAGGGCCAAAGAACTCTTAGACCTGTGGAA 963  
Db  
1486 GCCCAGAGCGTACAGCAGTGTCCGCACTGAGGGCCAAAGAACTCTTAGACCTGTGGAA 1545  
2Y  
964 AAGGCATTATGTGGGACCGCGCTGTGTGGCGGTGGATCCAGACCCGGGAGGTGTACTGT 1023  
Db  
1546 AGACACTATGCTTGGGACCTGCCCTCTAGCCTCTCAGCTTTGCAATGTCCCTTGTCTCG 1605  
2Y  
1024 ACGGACTGTGATGATCTTCTCTGTGTCAGCTGTGGGCTGTGCAATCCATGAAAATCTCAT 1083  
Db  
1606 ATGGACTGTGATGATCTTCTCTGTGTCAGCTGTGGGCTGTGCAATCCATGAAAATCTCAT 1665  
2Y  
1084 GAACCTCAGGGGAAAAGAAATTTAGAACAGCAGGCGCATGCTCTCATGGAATCTACA 1143  
Db  
1666 GATCCTCAGGGGAAAAGAACTTTAGAAATGAGACAGGCGCATGCTCTCATGGAATCTACA 1725  
2Y  
1144 GGGCCTGCAGGGCAATTTGCCCTCATTTGGTGGAGTCTGTCTTGTGAGGATCCAAATGTGC 1203  
Db  
1726 GGGCCCATGGGCGCTGCCCTCATCTGGCAGAACTGTACCTTGTGAGGATCCAAATGTGT 1785  
2Y  
1204 TACCGATGGCTGGCATCAGAAAGGATCTGTTCCCTGATCANGGAAAATGTGGCCCTGGGA 1263  
Db  
1786 CATCGCTGGCTTGCCTCGGAAGGATCTGTAATGTCTCATCATGGGAAATGTGGCCCTGGGA 1845  
2Y  
1264 CATCGTATTCTGAAGGCGCTCCAGAAATGACCGGGAGAGATCTATCAGGAGTCTT 1323  
Db  
1846 CATCGCATCTGAAGGCTGTCTGTCAGAAATGACGAGGAGAGAAATATCAGGGGCTCTC 1905  
2Y  
1324 TGCCAGTCTCCCTCTCTCTGAGAGAAATCTGTGAAATTCCTCTGCCAAATGGAATCT 1383  
Db  
1906 TGTCCAGTCTCTCTCTCTGAAACAAATGGCTTGTGAAATTCCTCTGCCAAATGGAATCT 1965  
2Y  
1384 GTGCTGAGCGATGGAGCGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1443  
Db  
1966 GTAGTGTGAGTGGAGCAGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2025  
2Y  
1444 GATGGGAAACAGACCCAGGTCAAGAACTTATCTCTGGCACTGGCTGGGAAAGGTGAAAGCCA 1503  
Db  
2026 GATGGGAAACAAACCCAGGTCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2085  
2Y  
1504 TGTCTCCCTAGTACGGCTCTCAGAGCATCTGTTGTGTAAAGCAATCTCTGTATGAGAG 1563  
Db  
2086 TGGCCTAGTATGAGGAACTCAGAGAGTACCGGCTATGCAATGATCATCTCTGTACACAG 2145  
2Y  
1564 CTTCACTGGGAGCATCGCTCTGGGCGCTTGTCTGAGGACATTTGTTAACTGCGCTT 1623  
Db  
2146 CTCTACTGGGAGCATCTGTTGGGCTTCTGTCTGTGAATTAATCTCTTGTGACTGTCTCTC 2205  
2Y  
1624 AATGCAACCAATGGCTGGAATGAGAGGCAACCGTGTGTGTAGGCATTCAGACTCGAGA 1683  
Db  
2206 AATGTGACCAATGGCTGGAATGAGAGGCAACATGTGTGGGTGGGCAATTCAGACACGAAAG 2265  
2Y  
1684 GTCTCTGTCTCAGAGTCACTGAGGAGTGAATGACCAAGAGTGTCCAGATCTACT 1743  
Db  
2266 GTCTCTGTCTCAGAGTCACTGAGGAGTGAATGACCAAGAGTGTCCAGAGTCTACA 2325  
2Y  
1744 CGACCTGAAACTGTGGCGGCTCTTTTCTCCCATGCAAAAAGACTGTATTTGACTGTCT 1803  
Db  
2326 CGGCCGAAACTGTGAAGACCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2385

1804 TTCACTGAGTGGACACACCTGCTGCCAAGGATGTCCAAGCAGGAATGCCACAGTAAACAG 1863  
Db  
2386 TTCACTGAGTGGACACACCTGCTGCCAAGGATGTCCAAGCAGGAATGTACATTAACAA 2445  
2Y  
1864 TCTCGATACAGAACTCATCTCAAGAGAGCCTCAATGGAGCCAGGAATGCCAGATACC 1923  
Db  
2446 TCTCGATACAGAACTCATCTCAAGAGAGCCTCAATGGAGTCAAGAGTGGCCAGACACC 2505  
2Y  
1924 TTATATAGGAGAGAGAGTGTGAAGATGTTTCTTGTGTCTCTGTATATCGGTGGAAGCCA 1983  
Db  
2506 TTGTTTGAAGAAAGAGAGTGTGAAGATATCTCACTGTGTCTCTCATCCGGTGAAGCCG 2565  
2Y  
1984 CAGAAATGGAGCCCTTGCATCTTAGTCCAGAGTGTGTCTGGCAGGAATTAACCGGCGAGC 2043  
Db  
2566 CAGAAATGGAGCTCTTGTATCTCTGTTCCAGAGTCTATCAGGACAGGGGAGAAACAGGCGACC 2625  
2Y  
2044 AGTGAAGCCTCTGGAAGGGGTACAAACAGAGCTCTCTCATGCATCTCTGTATGACAAAC 2103  
Db  
2626 AGTGAAGCCTCTGGAAGGGGTCTGAGAGAGCTCTCTCATGCATCTCTGTATGATTAAC 2685  
2Y  
2104 CGGTGACAGAGAAATGATGGAATGCTCAAGCAGACAAACCGCATGCTCTCTCTTGTGCAA 2163  
Db  
2686 CAGTCTGACAGAAATGACAGAGTGTCTCAAGCAGATGAATGSCATGCTCTCTCTTGTGCG 2745  
2Y  
2164 GAATGACAGTCCCATGTCTGAGAGAGTGTCACTCTCTCTCTCTCTCTCTCTCTCTCTCT 2223  
Db  
2746 GAATGACACCATTTCCCTCTCTGGGACGCTGTCACTCTCTCTCTCTCTCTCTCTCTCTCT 2805  
2Y  
2224 TGCTCCACGAACTGTGAAGCCACAAAAGTAGGGGGGACAGCTCACAGGGAAGAGCAGA 2283  
Db  
2806 TGCTCCAGAACTGTGAAGCAACTCAGATCAGGAGGCGGAGCTAAACAGGGAAGAGCAGG 2865  
2Y  
2284 AAGAGGAGAAATGCCAGGATTTGACCTTTTACCTCTAGTGGAGACAGAACTATGTCTCT 2343  
Db  
2866 AAGAGAGAGAAATGCCAGGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2925  
2Y  
2344 TGTGATGAATTTATATCCCACTTATGGAACCTGTGAGATTCATCTCTCTCTCTCTCTCTCT 2403  
Db  
2926 TGTGATGATTTATGTTCTATCTCTTATGGAACCTGTGAGCTTCTCTCTCTCTCTCTCTCT 2985  
2Y  
2404 AAGAGGAGCCTCACCC--GAGGACTCGGGTCAAGCAGACAGCAAGAAATGTGGAGAA 2460  
Db  
2986 AAGAGGATGCTCAGCAAGAGGAGTGTGGGTGCAAGAGATGACAAAGAAATGTGGAGAA 3045  
2Y  
2461 GGCCTGGCTTTCCAGCAGTAGCCTGTCTGATAAAATGGAAGACCTGTGACCCCTCC 2520  
Db  
3046 GGCCTGGCTTTCCAGCAATAGCATGTCTCAATATAAATGGAAGACCGTTGACCCCTCA 3105  
2Y  
2521 TTCTGACGACCTCTGGTTTACATTCAGAAAAATGTGTCTATTCCTCTGCCATTTGATTGC 2580  
Db  
3106 TTCTGCAACAGCTCTGGTTTACATTCAGAAAGCATGTGTCTATCCCTGTCTCTGTTGATTGC 3165  
2Y  
2581 AAGTTAAGCATTTGGTCTAGTTGGGGGTCTTGCAGTTCTCTCTCTCTCTCTCTCTCTCTCT 2640  
Db  
3166 AAGTTAAGTGTAGTGTCCAGCTGGGGTCTTGTGAGCTCATCTTGTGGGATTTGGAGTGAGA 3225  
2Y  
2641 ATTCGATCCAAATGGC 2656  
Db  
3226 ATTCGATCCAAATGGC 3241

## RESULT 3

AK047508

LOCUS

DEFINITION

AK047508 2266 bp mRNA linear HTC 20-SEP-2003  
Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length  
enriched library, clone:B930082A18 product:weakly similar to  
HYPOTHETICAL PROTEIN D1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo  
sapiens], full insert sequence.

ACCESSION

AK047508

VERSION

AK047508.1 GI:26092224

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)











Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 820)  
 NIH-MGC <http://mgc.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)  
 Seq primer: pYX-5.  
 Location/Qualifiers  
 1..820  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGS:6405956"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 12.5dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH-BMAP\_F00"  
 /note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lemmon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is TGAGAGACC. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

ORIGIN  
 Query Match 12.3%; Score 542.6; DB 13; Length 820;  
 Best Local Similarity 81.5%; Pred. No. 1.4e-156;  
 Matches 626; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 395 CAATGACGAGTCACCTTTAAACATCAAAAGTTACAAAGCACATCATTCATTCGAAGTCCT 454  
 DB 1 CAACGAGCAAGTCACCTTAAACATCAAGAGTTACAAAGCACACACCACTCCCGAGCTG 60  
 QY 455 GGGCAATGAGATAGTTATCAACACCGGAGGTTCTGTGTACAGAGTGATGGACAA 514  
 DB 61 GGGATGTAGTATAGGTTTCAACACCGGAGGTTGTGTGTACAGAGTGATGGAGAA 120  
 QY 515 ATGCTATGTTAAGCCTTTGCGCTTCAAGATTCCTCCCATTCGACCTGTTTCAGTCTGCATCA 574  
 DB 121 AGCCCTGTTAAGCCTTTGTTGTGTGATTCCTTCCTTCGACCTGTCAGCTTGTGTGA 180  
 QY 575 TGCCCAAGAGCTGTGAACCTCCAGTGGTCTCTCTGGAGCCCTGCTCCAGACATGCC 634  
 DB 181 TGCCAAAGAGCTGTGAGACTTCTGAATGGTCTCCATGAGCCCTGTTTCAAGACCTGCC 240  
 QY 635 GTTCAGGAGTCTCTTGCCAGATTTAGAGCAGAGCGCGGAGAGCTGAAGCACATGCTGA 694  
 DB 241 GTTCTGGAGACTTATCACTGTGAGTTAGAGCAGAGCGCGGAGAGCTGAAGCATTTGCTA 300  
 QY 695 TTGGAGGTGGAAAGAGGTGCTCTGAACTTCTTGAGAAAGAGCGCTGATGTTTGAAGGAG 754  
 DB 301 TTGGAGGTGGCCAGGATGATCCAGAACTTCTTGAGAAAGAGACCTGATGCTGAAGGGG 360

QY 755 AACTTCTGCGAGCATGTCCCGAGGTATTCCTGGAGAACTTCTGAATGGAAGAAATGCCAAG 814  
 DB 361 AATTCTTGCAACCATGTCCCGAGGTATTCCTGGAGAACTTCTGAATGGAAGAAATGCCAAG 420  
 QY 815 TCTCTCTCTCTCTCGAGCAGCAGCATCCCGACTGGCATGTGACGGGACCGCGTGTGTGGCG 874  
 DB 421 TCTCTCTCTCTCTCGAGCAGCAGCATCCATTGGCATGAGACAGCAATCTGTGTGTG 480  
 QY 875 GTGGATCCAGACCCCGGAGGTGTACTGTGTCAGAGCGGTACAGCAGCTGCCCATGA 934  
 DB 481 GTGGATCCAGACCCCGGAGGTGTACTGTGTCAGAGCGGTACAGCAGCAATTCATCAA 540  
 QY 935 GGGCCAAAGAGTCTCTAGACCTCTGGAAGAGGCAATATGTGTGGGACCGCCCGCTTGC 994  
 DB 541 GACCAAGAGTCTCCAGGCCCTGTGAAGGACATATGCTTGGGACCTGCCCTCAG 600  
 QY 995 CCTCTAGCTCTGCAATATCCCTTGTCTACGGACTGATAGTATCTTCTGTGTCAGCCT 1054  
 DB 601 CCTCTAGCTTTCGAATGTCTCTTGTCTCGATGAGTATCATGATGTCCTCTGTGTCACCT 660  
 QY 1055 GGGGCTGTGTCATCCATGAAGAACTGTGATGAACCTCAGGAGGAGGAGGATTTAGACGA 1114  
 DB 661 GNGGCCCATGTGTTTTTGTGAGAACTGCCATGATCTCAGGGGAGAGANAGGCTTTAGAATGA 720  
 QY 1115 GCGAGCGCATGTCTCTCATGAATCTACAGGCGCTCAGGCGCATTCGCC 1162  
 DB 721 GACAGCGCATGTACTCATGAGTCTACAGGCGCATTCAGGCGCGCTGCC 768

BY735628 717 bp mRNA linear EST 17-DEC-2002  
 BY735628 RIKEN full-length enriched, 1 cell embryo Mus musculus  
 CDNA clone I0C0030112 5', mRNA sequence.  
 BY735628  
 BY735628.1 GI:27148755  
 EST.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 717)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaudo, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I.,  
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C.,  
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
 Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V.,  
 Chothia, C., Corbani, D.E., Cousins, S., Della, E., Dragani, T.A.,  
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,  
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
 Numata, K., Okido, T., Pavan, W.J., Petrea, G., Pesole, G., Ramchandran, S.,  
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,  
 Ravasi, T., Reed, J.C., Reid, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
 Sandelin, A., Schneider, C., Sempile, C.A., Setou, M., Shimada, K.,  
 Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
 Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
 Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,  
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
 Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 JOURNAL  
 MEDLINE  
 22354683

PUBMED  
COMMENT

12466851  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9226  
Email: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, P.,  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,  
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,  
Kondo, S., Konno, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,  
Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N.,  
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,  
Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

## FEATURES

## source

1. 717  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="I0C0030112"  
/cell\_type="1 cell"  
/dev\_stage="1 cell embryo"  
/clone\_lib="RIKEN full-length enriched, 1 cell embryo"

## ORIGIN

Query Match 11.5%; Score 506.8; DB 13; Length 717;  
Best Local Similarity 84.7%; Pred. No. 1.9e-145;  
Matches 592; Conservative 0; Mismatches 103; Indels 4; Gaps 2;  
QY 2076 AGCTGCTCATGCTCTCTGATGACACCGGTGACGAGAAATGATGATGCTCAGCA 2135  
Db 14 AGCTGCTCATGCTCTCTGATGATACACCGGTGACGAGAAATGATGATGCTCAGCA 73  
QY 2136 GACAAAGCGATGCTCTCTGTCGAGAAATGACAGTCCCTGTCGAGAACTGCGAC 2195  
Db 74 GATGAATGGATGCTCTCTCTGTCGAGAAATGACAGTCCCTGTCGAGAACTGCGAC 133  
QY 2196 CTTTCACCTGTTGTTCAAGTTTACGCTGCTGTCAGCAAGTGTGAGCCACAAAAGTAG 2255  
Db 134 CTTTCACCTGTTGTTCAAGTTTACGCTGCTGTCAGCAAGTGTGAGCAACTCAGATCAG 193  
QY 2256 GCGGCGACAGCTTCACAGGGAAGCAAGAAAGGAGAAATGTCAGATTTCTGACCTTTA 2315  
Db 194 GAGCGCGAGCTTAACAGGGAAGCAAGAAAGGAGAAATGTCAGATTTCTGACCTTTA 253  
QY 2316 CCCTCTAGTGAGACAGAACTATGCTCTTGTGATGAATTTATATCCCACTTATGAAA 2375  
Db 254 CCCTCTAGTGAGAAATGAAACCATGCTCTGATACGTTTATGCTCTATGCTTATGAAA 313  
QY 2376 CTGCTCAGATTGCATTTCTTCCAGNAGCAGAGGAGCCTCACC---GAGACTGCGGGT 2432

Db 314 CTGCTCAGCTTGCTTCTTACCAGAGGCAAAAGGATGCTCAGCAAGGAGGATTTGGGT 373  
QY 2433 AAGAGCAGACAGCAAGAAATGTCGAGAGGCGCTGCGCTTTTCGAGCAGTAGCTGTTCTGA 2492  
Db 374 GCAAGGAGATGCAAAAGAAATGTCGAGAGGCGCTGCGCTTTTCGAGCAATAGATGCTCAA 433  
QY 2493 TAAAAATGGAAGACCTGTTGACCCCTCTCTGACAGCAGCTCTGGTTACATTCAGAAAA 2552  
Db 434 TATAAATGGAAGACCGGTGACCCCTCATTTCTGCAACAGCTCTGGTTACATTCAGGAAGC 493  
QY 2553 ATGTCATTCATCCCTGCGCCATTTGATTCGAAGTTAAGCGATGCTAGTTGGGGGCTTTG 2612  
Db 494 ATGTCATTCATCCCTGCGCCATTTGATTCGAAGTTAAGCGATGCTAGTTGGGGGCTTTG 553  
QY 2613 CAGTTTCATCTTTGTGGAATGGAAGTTCGAATTCGAATTCGAATTCGAATTCGAATTCGA 2672  
Db 554 CAGTCATCTTTGTGGATTTGAGTGGAGTTCGAATTCGAATTCGAATTCGAATTCGA 612  
QY 2673 CAATGGAGGAGCAGCATGTCCTCAAACTGGATCTCAAGAACTCAGTACATGAGGAGTCCC 2732  
Db 613 CAGTGGAGGTGCGCCATGTCCTCCANACTGGATCTCAAGAACTCAGTTCATGAGGAGTCCC 672  
QY 2733 ATGTTACAGTCAAGTCAATCAGTATTCCTGGGTTGTAGA 2771  
Db 673 CATGTACAGTCAAGTCAATCAGTATTCCTGGGTTGTAGA 711

## RESULT 10

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

D	b		361	TGTCCTCGTGAACGTTTAATGTCTCATCTCTAATGGAAACTGTCAGCTTGCACTTACCA	420
Q	y		2398	GAAGGCAGAAGGAGGAGCCTCACCC---GAGGACTGCGGGTACAAGCAGACAGCAAGAATGT	2454
D	b		421	GAAGGCCAAAAGGATGCTCAGCAAGGAGGATTGTGGTGCAAGGAGATGACAAGAATGT	480
Q	y		2455	GGAGAAGGCTCGGCTTTTCGAGCAGTAGCCTGTCTGATAAAATGGAAGACCTGTTGAC	2514
D	b		481	GGAGAAGGCTCGGCTTTTCGAGCAA'TAGCATGCTCCAATATAATGAAGACCGGTTGAC	540
Q	y		2515	CCTCCTCTTCGAGCAGCTCTCGTTACATTCAGAAAAAATGTCTATTCCTGCCCATTT	2574
D	b		541	CCTCATTTGCAACAGCTCTGGTTACATTCAGGAAGCATGTGTCATCCCTGTCGGTTT	600
Q	y		2575	GATTGCAAGTTAAGCAATGTGTTAGTTGGGGTCTTTGCAGTTTCATCTTGTGGAATTGGA	2634
D	b		601	GATTGCAAGCTAAGTCACTGGTCCAGCTGGGTTCTTGCAGCTCATCTTGTGGGAITGGA	660
Q	y		2635	GTGAGAATTCGATCCAAATGGC	2656
D	b		661	GTGAGAATTCGATCAAAATGGC	682
R	E	S	U	L	T
RESULT 11					
BI988413					
LOCUS					
DEFINITION					
ACCESION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
COMMENT					
FEATURES					
source					
ORIGIN					
Query Match					
Best Local Similarity					
Matches					
Q	y		3065	GGACTTGTGTTGAAGACTCAAGTGCAGCGCTTCCTCCCTGAATGAAAAATGCTTCAGT	3124
D	b		1	GAACTTTGTTGGGAAGACTCGCAGGTGGGGCCCTCCCTCAATGAAAACCTGTTTCAGT	60
Q	y		3125	TCCAGTACATCTAACAGAGTGGAGACATGCCAGCTGAGTGA AAAACGACCCCTGTGTC	3184
D	b		61	TCCAGTACATCTGACAGAGTGGAGACATGCCAGCTGAGTGA AAAATGTTTCCTGTGCCC	120
Q	y		3185	AAGCGCTCAGGACCCGCTGCTAAGCTGTGTGTGTCAGTGTATGGCAACCCAGTCAAGATGG	3244

121	AGG	GTG	TGAG	ACCG	CCCT	CTCT	GAGT	GTGT	GGC	GAT	GTG	CAAGT	CTCT	CAG	CTG	180
3245	ACCA	ATGT	GAG	CAG	CA	TAA	TTT	TG	GA	AG	CC	CC	CA	GA	AT	3304
181	ACC	ACT	CG	AG	CAG	CGT	TAA	CT	CG	AA	AA	AA	CC	CA	AG	240
3305	AAT	CGT	GGT	CA	ACT	GT	CA	CT	CA	GGT	TG	AG	CG	CTT	TG	3364
241	AAT	GT	GGT	CA	ACT	GT	CA	CT	CA	GGT	TG	AG	CG	CTT	TG	300
3365	GTG	CCAT	TG	AG	GT	CG	CA	AT	GAG	CG	CG	AC	T	CG	AT	3424
301	GTG	CC	AA	AG	AG	GT	CG	CA	T	AG	CG	CA	CT	CG	AT	360
3425	GAC	GC	CA	TG	CC	CA	AG	AG	CTT	TAC	CC	AG	GA	AA	AA	3484
361	GTG	CA	CA	AT	GC	CC	CA	GA	AA	CT	CAC	CA	G	AG	AA	420
3485	GCT	GGT	CC	T	GG	CA	ACT	GT	CG	CA	T	GA	AT	T	GA	3544
421	GTG	GGT	CC	T	GG	CA	ACT	GT	CG	CA	T	GA	AT	T	GA	480
3545	TTC	AG	AT	CCG	CAG	CG	CTT	TC	TG	CA	TG	TC	CA	AG	TG	3604
481	TCC	AG	GT	CG	CAG	CTT	TC	CT	GT	GT	AG	TCC	CA	AT	GC	540
3605	GTG	TC	GAG	AT	GC	AT	CT	GT	TG	GA	AA	T	GC	CTT	TG	3664
541	CTG	TG	AG	AG	CG	CC	CT	GT	TG	GA	AA	T	GC	CTT	TG	600

RESULT 12

AV590588

LOCUS

DEFINITION

5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

565 bp

mRNA

linear

EST 27-NOV-2001

Bos taurus brain fetus

Bos taurus cDNA clone

E18013E03

1

(bases 1 to 565)

Takasuga, A., Hirotsune, S., Itoh, R., Jitohzono, A., Suzuki, H., Aso, H., and Sugimoto, Y.

Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs

Nucleic Acids Res. 29 (22), E108 (2001)

21570554

11713328

Contact: Yoshikazu Sugimoto

Animal Genetics Division

Shirakawa Institute of Animal Genetics

Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan

Tel: 81-248-25-5641

Fax: 81-248-25-5725

Email: kazu@ococa.ocn.ne.jp

Single pass sequencing.

This clone was obtained from a polyA-deleted cDNA library.

Location/Qualifiers

1..565

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

/clone="E18013E03"

/tissue\_type="brain"

/dev\_stage="fetus"

/lab\_host="DH10B"

/clone\_lib="Bos taurus brain fetus"





Y 1970 ATCGGTGGAGCCACAGAAATGGAGCCCTTGTGATCTTAGTCCAGAGTCTCTGGCAGG 2029  
b 421 ATCGATGGAGGCCACAGAAATGGAGCCCTTGTGATCTTAGTCCAGAGTCTCTGGCAGG 480  
Y 2030 GAATAACGGGACAGTGTGAAGCCTGTGGAAGGGGTTTAAACACAGAGCTGTCTCATGCA 2089  
b 481 GACTAATAGGCACAGTGAACCTTGGGAAGGGAGTGCAATAAGAGCTGTCTCATGCA 540  
Y 2090 TCTCTGATCA 2099  
b 541 TCTCTGATAA 550

RESULT 15  
LOCUS BI991468 600 bp mRNA linear EST 20-DEC-2001  
DEFINITION 4101-38 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,  
mRNA sequence.  
ACCESSION BI991468  
VERSION BI991468.1 GI:17962491  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 600)  
AUTHORS Mu, X., Zhao, S., Pershad, R., Hsieh, T.-P., Scarpa, A., Wang, S.W.,  
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.  
TITLE Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
MEDLINE 21671825  
PUBMED 11812828  
COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.  
Location/Qualifiers  
1. .600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/clone\_lib="Mouse E14.5 retina lambda:ZAP II Library"

FEATURES  
source

Query Match 9.5%; Score 422.4; DB 12; Length 600;  
Best Local Similarity 82.7%; Pred. No. 3.5e-119;  
Matches 496; Conservative 0; Mismatches 101; Indels 3; Gaps 1;  
Y 1882 ATCCAGAAGCAGCAATGGAGGCCAGGAATGCCAGATACCTTTATATAGGAGAGAG 1941  
b 1 ATCCAGGAGCAGCAATGGGGTCAAGAGTGCCACACACCTTTGTTGAAGAAGAGAG 60  
Y 1942 TGTGAAGATGTTCTCTGTGTCCTGTATATCGGTGGAAGCCACAGAAATGGAGCCCTTGC 2001  
b 61 TGTGAAGATATCTCACTGTGTCTTATACCGGTGGAAGCCGAGAAATGGAGCTTTGT 120  
Y 2002 ATCTTAGTGCCAGAGTCTGTGCGGGAATTAACGGGCGAGGTGAAGCCTGTGGAAG 2061  
b 121 ATCTGTGTTCCAGAGTCTATCAGCAGGGGAGAACAGCCACCATGAGGATGTGGAG 180  
Y 2062 GGGTTACAAACAGAGTGTTCTATGATCTCTGATCAACCGGTGAGCAAAATGATG 2121  
b 181 GGGCTGCAGACAGAGTGTCTCATGATCTCTGATGATTAACCATCTGAGAAATGACA 240  
Y 2122 GAATGCTCCAGCAGACAGCAACCGCATGCTCTCTGTCGAAGATGCACAGTCCCATGT 2181  
b 241 GAGTGCCTCAAGCAGATGATGGCATGCTCTCCCTTTGTGAGGAATGCCATCTCCTGT 300

QY 2182 CGAAGAGACTGCACCTTCACTGCTTGGTCCAAAGTTTACGCCCTGTCTCCACGAACCTGAA 2241  
Db 301 CGGAGCAGACTGCACCTTCACTGCTTGGTCTAAGTTCACTCCCTGTCTCCAGAACTGTGAA 360  
QY 2242 GCCACAAAAGTAGTAGGGCGGACAGCTCAAGGGAAAAGCAGAAAGAGGAATGCCAG 2301  
Db 361 GCAACTCAGATCAGAGGGCGGCGAGCTTAACAGGGAAAAGCAGGAAGAGAGAGTCCAG 420  
QY 2302 GATTCGTACCTTTTACCTCTAGTGGAGACAGAACTATGTCTTGTGATGAATTTATATCC 2361  
Db 421 GATGCGTCCCTGTACCTCTGGTGGAAATGGAACCATGTCCCTGTGTACGTTTATGTCT 480  
QY 2362 CAACCTTATGGAAGTGGTTCAGATTGCATTTCTCCAGAAAGCAGAAAGGAGCCTCACC-- 2419  
Db 481 CATCCTTATGGAAGTGGTTCAGATTGCATTTCTACAGAAAGCAGAAAGGAGTGTCTCAGCA 540  
QY 2420 -GAGGACTGCGGTACAAGCAGACAGCAAGAAATGTGGAAAGCCTTGGCTTTTCGAGCA 2478  
Db 541 GGAGGATTTGGGTGCAAGGAGATGCAAGAATGTGGAGAGGCGTGCCTTTCCAGCA 600

Search completed: February 23, 2004, 20:29:32  
Job time : 10056.4 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

DM nucleic - nucleic search, using sw model

run on: February 23, 2004, 12:03:01 ; Search time 329.257 Seconds  
(without alignments)  
7412.668 Million cell updates/sec

Title: US-10-022-710-1  
Perfect score: 4398  
Sequence: 1 atgggtgcgtgattcagaa.....atggagacttagacatgtaa 4398

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/prodata/2/ina/5A.COMB.seq\*
  - 2: /cgn2\_6/prodata/2/ina/5B.COMB.seq\*
  - 3: /cgn2\_6/prodata/2/ina/6A.COMB.seq\*
  - 4: /cgn2\_6/prodata/2/ina/6B.COMB.seq\*
  - 5: /cgn2\_6/prodata/2/ina/PCTUS.COMB.seq\*
  - 6: /cgn2\_6/prodata/2/ina/backfileseq1.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1229.6	28.0	6378	4	US-09-620-312D-332
2	161.8	3.7	579	4	Sequence 332, Appl
3	44	1.0	3226	1	US-07-862-021B-11
4	44	1.0	3226	1	US-08-313-288B-11
5	44	1.0	3226	5	PCT-US93-03164-11
6	42.8	1.0	7218	1	US-08-232-463-14
7	39.8	0.9	4061	4	US-09-132-769-4
8	39.2	0.9	3079	4	US-09-132-769-2
9	39.2	0.9	3079	4	US-09-640-173-184
10	39.2	0.9	3079	4	US-09-713-550-184
11	39.2	0.9	3999	4	US-09-640-173-177
12	39.2	0.9	3999	4	US-09-713-550-177
13	39.2	0.9	4866	4	US-09-976-594-353
14	37	0.8	505	4	US-09-621-976-15639
15	37	0.8	576	4	US-09-397-992A-25
16	37	0.8	576	4	US-09-971-843-25
17	37	0.8	832	4	US-09-621-976-2813
18	36.6	0.8	832	4	US-09-621-976-2813
19	36.6	0.8	4029	1	US-07-862-021B-9
20	36.6	0.8	4029	1	US-08-313-288B-9
21	36.6	0.8	4029	4	US-09-132-769-6
22	36.6	0.8	4029	5	PCT-US93-03164-9
23	36.2	0.8	882	4	US-09-252-991A-2427
24	36.2	0.8	993	4	US-09-252-991A-2588
25	35.8	0.8	474	4	US-09-621-976-18033
26	35	0.8	1668	4	US-09-252-991A-10888
27	35	0.8	1848	4	US-09-252-991A-10829

Sequence 2427, Ap  
Sequence 3, Appli  
Sequence 15, Appl  
Sequence 19, Appl  
Sequence 5, Appli  
Sequence 165, App  
Sequence 298, App  
Sequence 50, Appl  
Sequence 67, Appl  
Sequence 67, Appl  
Sequence 67, Appl  
Sequence 67, Appl  
Sequence 7862, Ap  
Sequence 7895, Ap  
Sequence 7515, Ap  
Sequence 7823, Ap  
Sequence 8976, Ap

ALIGNMENTS

RESULT 1

US-09-620-312D-332  
; Sequence 332 Application US/09620312D  
; Patent No. 6569862  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhai, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yundong  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. 556962e1 Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 332  
; LENGTH: 6378  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (271)..(5037)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(6378)  
; OTHER INFORMATION: n = a,t,c or g  
US-09-620-312D-332

Query Match 28.0%; Score 1229.6; DB 4; Length 6378;  
Best Local Similarity 57.6%; Pred.No. 0;  
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;

2Y 10 TGCAATCAGAAAGCTGAACCGAACTGTGGTTGCAAAATGCAAAATATGCAACACACTTGGCCCTT 69  
Db 541 TGCATCAGAAAGACAAAGACATCTCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 600  
2Y 70 CAGGCTCTACAGAAAGGCTTGGCTCAATCTCTTGTTCCTGGGATTTGTAGTATCTGAG 129  
Db 601 AAGCCTCTCTGAGCAGGCTTGGCTCAATCTCTTGTTCCTGGGATTTGTAGTATCTGAG 660  
2Y 130 TTTCTTACCATGTTCCCACTCTAGCAAGGATGTGGGAAGAAATGCAAGATATGCAATCTGCG 189  
Db 661 TTTTCTGCTGTTCCGAATGCTCCAAGACCTGCGGAGCGGCTCCAGCACCGAGCGCT 720  
2Y 190 GCGGTCAATGATCCCTCTCTTGTGGTGGTTGCAATGTCCTCAATCTGACTGATCAAGA 249  
Db 721 CATGTGTGGCCCGCCGAGTTTGGAGGCTTGGCTGTCCTCAAACTGACGGAGTTCCAG 780  
2Y 250 GCCTGTGATGTCCTCAATCTCTGCTCTTCTTGGGAGAGGAATATACATTTAGCCTTAAG 309  
Db 781 GTGTG-----CCAAATCCAGTCCATGCGAGCGGAGGCTCAGGTACACCTGCAAT 831  
2Y 310 GTTGGACATGAGTAATGCACTGCTCTCATCTTTAAAGAAATTTAATCCAGCGGAAGA 369  
Db 832 GTGGGCGCTTGGAGCACTGCTCAATGCCCACTCTCCGACAAAGTAAGACAAAGGAGA 891  
2Y 370 ACTGTTCTGGAATTTAA-----CTCTGATTCAAATGAG 402  
Db 892 CGCGGAGAAATAAGAAACGGGAAAGGACCGGACGAAAGAGTAAGGATCCAGAGCC 951  
2Y 403 CGAGTCACCTTTAAACATCAAGTTACA-----AAGCACATCATCATTCGAAGTCT 453  
Db 952 CGCGAGCTTTAAGAAAGAAAGAGAAACAGAAACAGGACGAGCAACAGAAAGAAATAT 1011  
2Y 454 TGGCAATAGAGATAGTTATCAAAACCGGAGGTTCTGTTCAAGAAAGTATGACAA 513  
Db 1012 TGGACATCCAGATTTGATATCAGACAGAGAGGTTATGTCATTAACAAAGACGGGAAA 1071  
2Y 514 AATGCTATGTTAAGCCTTTGCTTTCAAGATTCCTTCCCATTTGACTGTTTCACTCTGCAATC 573  
Db 1072 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGAGCTTCCAAATGACCTTCCAGTCTCTGTG 1131  
2Y 574 ATGCCAAAGACTGTGAACCTCCAGTGTCTCTCGAGGCTTCTGAGAGCCCTCTCCAGACATGC 633  
Db 1132 ATCAACAAAGAGTCCAGGTTTCCGAGTGTGTCAGAGTGGAGCCCTCTCTCAAAACATGC 1191  
2Y 634 CGTTCAGGAGTCTCTTCCAGGATTTAGGACAGGAGCCGGAACGTGAAGACATGGCT 693  
Db 1192 CATGACATGGTGTCCCTGAGGCACTCTGTAGGACACGAACCATCAGGAGTTCCC 1251  
2Y 694 ATGAGGTGGAAGAGGTGCTGAACTTCTTGAGAAAGAGGCTCTGATTTGTTGAAGGA 753  
Db 1252 ATTGSCAGTGAAGAGGTGCTCCAGATTTGAAGAAAGAAACCTCTGTTCTCAAGGA 1311  
2Y 754 GAACCTTCTGCAATGTCCTCAGGATTTCTTGGGAATCTCTGAATGGAAGATGCAA 813  
Db 1312 GATGAGTTTCTCTCTGACAGATGTCCTGAGCACTGTCAGAGTGAATGAGTGGCT 1371  
2Y 814 GTCTCTCTCTCTCTGACAGCAGGATCCCACTGGCATGTGACGGGACCGGTGTGGC 873  
Db 1372 GTGGACCTTTGCTCAGTCAGCAGCAAGAGGCGCGCAACCAAGACCGCCCTCTGTGA 1431  
2Y 874 GTGGGATCCAGACCGGAGGTGACTGTGCCAG-----AGGTACCAAGCA 921  
Db 1432 GGGGGATCCAGACCGGAGGTGACTGTGCGTGGAGGCAACGAAACCTCTCTCAAA 1491  
2Y 922 GCTGCGCACTCAGGCGCAAGGAGTCTCTAGACCTGTGGAAGGCAATATGTGTGGGA 981  
Db 1492 TTAAGTACCCACAGAAACAAAGAGCTTCAAGGCAATGAGACTTTAAATATATGCACTGA 1551  
2Y 982 CCGGCGGCTTCTCTCTCTGATCTGATATCTCTGATATCTCTGATGATCTCTGATGATCT 1041  
Db 1552 CCTATCCCTTAATACATACAGCTGTGCCACATCTCTTGTCCAACTGAATGTGAAGTTCA 1611  
2Y 1042 TCCTGTGAGCGCTGGGCGCTGTGCATCCATGAAACTGTGCATGAACTCTCAGGAGGAAAAA 1101

Db 1612 CTTTGGTCAGCTTGGGACCTTGTACTTATGAAACTGTAAATGATCAGAGGAAAAA 1671  
QY 1102 GATTTAGAACGAGGACG-----CCATGCTCTCATGGAATCTACAGGCGCTTGA 1152  
Db 1672 GCTTCAAACTGAGAGAGCGGCAATACCAATGAGCCACTGGAGCTCTGGGTAAAC 1731  
QY 1153 GGGCAATGCCCTCAATTTGGTGGAGTCTGTTCTTGTGAGGATCCAAATGTTACCGATGG 1212  
Db 1732 GGAACCTGCCCTCACTTACTGGAAGCAATTTCCCTGTGAAGAGCCTGCTGTATGACTGG 1791  
QY 1213 CTGCA---TCAGAAGGATCTGTTTCCCTGATCATGAA---AATGTGGCCTGGGACAT 1266  
Db 1792 AAGCAGTGAAGCTGAGAGACTGAGGAGACTGAGCCAGATTAACGAAGAGTGTGTCCAGGACG 1851  
QY 1267 CGTATTTGAAGGCGCTGTCAGAAATGACCGGAGAAAGATGATCAGGAGTCTTTG 1326  
Db 1852 CAAGTTCAAGAGGTTGTGTGATCAACAGTGTGAGAGAAAGTTGACAGACAGCTGTGC 1911  
QY 1327 CCAGTTCCCTCTCTGAGAGGAGTCTTGTGAAATTCCTCCGAAATGAGCTGTGTG 1386  
Db 1912 AGAGATGCCATCTTCCCATCTCTGTCCTGTCACACACCTCTCAGGGAACACAGAA 2031  
QY 1387 CTGAGCAGTGGAGGAGTGTCTATCTTCTTCCAGTCTTCTTCAAAATAAAATCAAGAT 1446  
Db 1972 CTCACACATGCTCTACGTGTCTCTCTCTCTCAGACCTCTCAGGGAACACAGAA 2031  
QY 1447 GGGAAACAGACAGGTCAAGAACTATCTCTGCG---ACTGGCTGGGAGGTTGGAAGCCA 1503  
Db 2032 GGGAAACAGATACAGAGCAGATCTCTGCGCTATGCGGCTGAGAGGTTGGAATTCG 2091  
QY 1504 TGTCTCCCTTAGTCAGGCTCTCCAGAGCATCTTGTGTGTAATGACCATTCCTGTATGAG 1563  
Db 2092 TGTCAAAATGACAGTCTTTCAGAGATGACGAGTGTATGAGTATCTTGCACAGTG 2151  
QY 1564 CTTCTCTGGAGACATCCCTTGGGCGCTTGTCTGAGGACATCTGTAATGCGCTT 1623  
Db 2152 TACCACTGGCAACTGCTCTTGGGCGAGTGTGAGGACCTCAGTATGCTCTCTT 2211  
QY 1624 AATGCAACCTTGGCTGGAATGGAGAACCCAGCTGTGTGTAGGCAATTCAGACTCGGAGA 1683  
Db 2212 AACACAACTACAGTGTGAATGGGAGGCTCTCTCTGTCGCAATGACAGAGAA 2271  
QY 1684 GTCTCTGTCTCAAGAGTCACTGGGCAAGATTAATGACCAAAAGATGTCAGATTTCTACT 1743  
Db 2272 GTCACTGTGTGCGAGTCAATGTGGGCAAGTGGGACCCCAAAATGTCCTGAAAGCCTT 2331  
QY 1744 CGACCTGAACTGTGGCGCTCTCTCTCTCCATGCAAAAGAGCTGTATTTGACTGCT 1803  
Db 2332 CGACCTGAACTGTGAAGCCTTGTCTGTTCTTGTGTAAGAGGACTGTATTTGACCCCA 2391  
QY 1804 TTCAGTGTGACACCTCTGCGCAAGGATGTCCCAAGCAGGAAATGCCACAGT---AAAA 1860  
Db 2392 TATAGTACTGAGACATATGCGCTCTTCTGTAAAGAGGAGCTCCAGATATCAGGAAG 2451  
QY 1861 CAGTCTCGATACAGATCATCTCAAGAGCAGCCATGAGGCGGAGGATGCGGAGAT 1920  
Db 2452 CAGTCTTAGGATGAGGCTCATCTTCACTGCGCAGCAACGGGCGGAGACTGACAGAT 2511  
QY 1921 ACTTATATGAGGAGAGAGTGTGAAGATGTTCTTGTCTGTATATGCTGTGGAAG 1980  
Db 2512 CCCCTCTATGAAGAGAGGCTGTGAGGACCTCAAGCGTGCAGGCTACAGGTGGAAG 2571  
QY 1981 CCACAGAAATGAGCGCTTGGATCTTAGTGCCAGAGTCTGTCTGCGAGGAAATACGGGC 2040  
Db 2572 ACTCAAAATGCGCAGATGCGCAATTAGTCTCTGAGCGTGCAGCAACAGACAGCCCTGGA 2631  
QY 2041 AGCAGTGAAGCCTGTGGAAGGGGTTACAAACAGAGCTGTCTCATGATCTCTCATGAC 2100  
Db 2632 GCACAGAAAGGTTGTGGGCTTGGGCGACAGGCAAGGCCATTTACTTTGCGAAGCAAGAT 2691  
QY 2101 AACCGTCAAGCAAAATGATGAAATGCTCAAGCAGAGCAAAACGGCATGCTCTCTCTGTG 2160

2692 GGAGGACAGGCTGGAATCCATGAGTGCCTACAGTATGCAGGCCCTGTGCAGCCCTTACC 2751  
2161 CAAAGATGACAGTCCCATGTGAGAGACTGACCTTCACTGCTTGTGTCNAGTTTACG 2220  
2752 CAGGCTGCAGATCCCTGCCAGGATGACTGCTCAATTTACAGAGTGGTCCAAGTTTCT 2811  
2221 CCTGTCTCCAGCAACTGTGAAGCCCAAAAGTAGCGGCGACAGCTCACAGGGAAGAC 2280  
2812 TCATGCAATGGAGCTGTGGTGCAGTTAGGACAGAAAGCGCACTCTTGTGGAAGAAAGT 2871  
2281 AGAAGAAGGAAGATGCCAGAGTTCTGACCTTTACCTCTAGTGGAGACAGACTATGT 2340  
2872 AAAAAAAGAAAAATGTAATAATTTCCATTTGTATCCCTGATTTGAGACTCAGTATGT 2931  
2341 CCTGTGATGAATTTATATCCCAACCTTATGAAACTGGTCAAGTTGCAATCTCCAGAA 2400  
2932 CTTTGTGACAAATATATGACACAACTGTGGGAAGTGTGAGACTGTATTTTACAGAG 2991  
2401 GGCAGAGGAGGCTCACCGAGACTGCGGGTACAAGGACAGAGACAGAAAGAAATGTGGAA 2460  
2992 GGAAGAGTGGAGTGTGCTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGGACAA 3051  
2461 GGCCTGGCTTTCGACAGTACCTGCTCTGATATAAATGGAAGACCTGTTGACCCCTCC 2520  
3052 GATATCGTTACCAAGCAATGGCATCTAGATCAAAATGGCAGGCTTGTGGAACAATCT 3111  
2521 TTCTGCAGCAGCTCTGGTTACATTCAGAAATAATGTCTCAATCCCTGCCCATTTGATTGC 2580  
3112 AGATGTAACAGCCATGTTTACATTTAGGAGGCGCTGCATCATCCCTGCCCTCAGACTGC 3171  
2581 AAGTTAAGCAATGGTCTAGTTGGGGCTTTCAGTTCACTTGTGGAAATGGAGTGAGA 2640  
3172 AAGCTAGTGAAGTGGTCCAACTGCTGCGCTGAGCAAGTCTGTGGGAGTGGTGTGAAG 3231  
2641 ATTCTGATCCAAATGGCTTAAAGAAACCTTCAATGGAGGACGACCATGTCCCAAACTG 2700  
3232 GTTCGTTCTAAATGGCTGGTGAAGAACCATATATGGAGGAGGCTTGGCCCAACTG 3291  
2701 GATC-----TCAGAAATCAGGTACATGAGGAGTCCCATGTTACAGTGAAGTCAATCAG 2754  
3292 GACCATGTCAACAGGACAGAGTGTATGAGTTGTCCCATGCCACAGTGAATGCAACCCAG 3351  
2755 TATTCTGGGTTGTAGAACACTGGTCTTCATGCAAAATCAAAATCAGCTGAGTGCCTCTG 2814  
3352 TACTATGGGTACAGAGCCCTGAGCACTGCAAGGTGACCTTGTGAATATGCGGAG 3411  
2815 CGCTGTGGAGGAGAAACAATCTAGAAATACAGATGTGT---GAATACGCGGATGT 2871  
3412 AACTGTGGAGAGGCGGTGCAAAACCGAAAGTGAATGATGCATGCAGAAATACAGCAGATGC 3471  
2872 GAAGTGGAGCAGTGGATACCACTGTGCAACAGGATGAAATCCCCAGAAACCCAG 2931  
3472 CCTTCTGAACATGTAGAGGATTTACTCTGTGACCCAGAGAGATGCCCTGGCTCTAGA 3531  
2932 TCCTGTTCTCTTATGTCTCCCAATGAGTGTGTCATGTCTGAGTGGGAGCTTTGGAGCAA 2991  
3532 GTGTGCAAAATACCATCCCTGAGGACTGTGTGATATCTGAATGGGTCCATGGACCCAA 3591  
2992 TGCCACAGTCACTGCGATCCCAACAAAT---GCAGAGAAAGCTGCCACCTGCTTAAGA 3048  
3592 TGTGTTTTGGCTTGCATCAAGAGAGTTTCCGGCAAGGTGAGCTGATCCCATCAGACAA 3651  
3049 CCATCACTGAATCAAGGACTGTGTGTAAGACTCACAGGTGCAGGCTTGCCTCTCTGAAT 3108  
3652 CCAGCTGATGAGGAAGATCTTGGCTTAATGCTTTGAGAAAGAACCTGTAACCTGAC 3711  
3109 GAAATTTGCTTCCAGTTCAGTCAATCTAACAGAGTGGAGCACTGCCAGCTGAGTGAA 3168  
3712 AAAAATGCTACCACTATGATTAATATGTAACAGACTGGAGTACATGTACAGTGAAGT 3771  
3169 AACGCACCTGTGTGTAAGGCTCAGAGCCGCTGCTAAGCTGTGTGTCAGTGAATGCC 3228  
3772 AAGCGAGTTTGTGGAATGGNAATAAACAAGGATGTGGATTTGATTTGATGATGCC 3931

QY 3229 AAGCAGTCAGCTGACCAATCTGTAGCAGCATAAATTTGAGAAAGCCCCAGAGAATGAGC 3288  
Db 3832 AAGTCAGTTGACCTGAAATATTTGTGAGCGCTTGGCTTGGAGAGAACTGGCAGATGAC 3891  
QY 3289 ATTCCCTGCTGTGTGAAATCGGTGGTCACTGTCACTGTCAAGGTGAGCGGTGGACA 3348  
Db 3892 ACGTCTGCATGTGTGAAATGCCCTGTGAACTGTGAGCTTCTGATTTGCTCTCTGTC 3951  
QY 3349 GAGTGTTCACAGACCTGTGGCCATGGAGTCGAATGAGCCGAGCTCGATTTATCATTTATG 3408  
Db 3952 GATGTTTCTCAAACATGTGCTTCAAGGAAATATATCCGAAGACGACAGTGCACCCAG 4011  
QY 3409 CCACCCCAAGGAGAGACGCGCCATGCCACAGAGCTTACCACAGAGAAACCTGCGCCA 3468  
Db 4012 CCCTTTCAAGGTGATGGAACCATGCCCTTCCCTGATGACCAAGTCCAAACCCCTGCCCA 4071  
QY 3469 GTGACCCCTCTCTACAGCTGGTCTTGGCAATGGTCTGCAATGTAATTTGAGGGTGGGA 3528  
Db 4072 GTGAAGCCTTGTATCGGTGGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGAGGCC 4131  
QY 3529 GACTGTGGGGAAGAGTTCAGATCCGAGCCTTTCCTGCAATGGTCCACAGTGGTTCATA 3588  
Db 4132 CAGTGTGAGAGAGGACCAAGAACATTTCTTGTAGTAAAGTATGGTCAAGCT 4191  
QY 3589 TCTCATGCACTGGAAGTGTGAGGATGCACTGTGTGGAGAAATGCTGCTGACATTTGAATCAT 3638  
Db 4192 GATGATTTCAAGAAAGTGGTGGATGAGGAATCTGTGCTGACATTTGAATCATTTATGAT 4251  
QY 3639 -----TCAGGACAGCATCTTGAAGCAGCTGTGTCTGTGCTTGGCCCTTGGCCAGGAGCTGCCAT 3693  
Db 4252 GGTAAATAAATAATGTTCTTGGAGGAATCTTGCAGCCAGCTTGGCCAGTGTAT 4311  
QY 3694 TTAACAGAAATGGTCAGAGTGGAGACATATGTAATTAACCTGCAATTTGATGGAAGAGCTTT 3753  
Db 4312 TTGAAGGACTGTTCTTCTGAGGCTGTGTGAGTCACTGTGTGATGTGTGAGGATCTA 4371  
QY 3754 GAGACTGTGGGCCCGCAGTCTAGATCAAGGACTTTTATAATTCAGTCTTTTGAAGAACAA 3813  
Db 4372 GCTTTGTGTAATACAGGTCAAGTCCAGACCGGTGATTTATACAGAACTAGAGANTCAG 4431  
QY 3814 GACAGCTGCCCCCAACAGGTTCTAGAAACAGCCCTTGTACAGGAGCAAAATGTTATCAC 3873  
Db 4432 CATCTGTGCCCAGAGCAGATGTTAGAAACAAATCATGTTATGATGACAGCTGTATGAA 4491  
QY 3874 TACACATGGAAGCAAGTCTTTGGAACAATAACGAACTGTATGTCGCCAGCTTCA 3933  
Db 4492 TATAATGGAATGGCCAGTGTGTGAAGGGCTCTTCCGAAACAGTGTGTGTCAAGTCA 4551  
QY 3934 GATGGCGTTAATGTCAAGAGGAGTGTCTCCCTCAGGCCGCTCTGTGCTGCAATTCGGCAG 3993  
Db 4552 GATGTTAATGTAACAGGGGGCTGCTTGGTGTATGAGCCAGCTGATGCCAGGCTCT 4611  
QY 3994 TGCATTCAGCTTCAGAAACCTTTCTCTACTGTACACAGGTTGAGTCTGTGGTGT 4053  
Db 4612 TGTAAACCCACCGTGTAGTCAACCCCTCTGCTGTAGCAGACAAACCAATGCAATGT 4671  
QY 4054 GAGAAGGCTATACAGAGATAATGAAATCAAAATGTTTCTCTGGAATTAAGTCA----- 4105  
Db 4672 GAAGAGGTTACATGAGTCAATCTTCTTAACAGCACCTTGTAGCAATGACACTTATC 4731  
QY 4106 -----TGAAAGTACAGGCTCAGAGATAAAGAGTGAATGTGAAAGAACTTTCTG-G 4157  
Db 4732 CCGCTGGTGTGTTTACCCACCATGGAGGACAAAGAGGAGATGTGAAACCCAGTCGGCT 4791  
QY 4158 GAAAGACAGCTGTGAAATTCAAATAATACATGATATTTTAAAGGA--TGCTCTCTCAA 4215  
Db 4792 GTACATCCAAACCCACCTTCCAGTAAACCCAGAGAGCGGGAAGGACCTGTTTCTACAG 4851  
QY 4216 CCATTTGATCCAGATGGCCGAGTAAAAATTTGGGTTTATGGCGTTTTCAGTGGCGCTTTT 4275  
Db 4852 CCATTTGGCCAGATGGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGCGCATTT 4911

QY 4276 CTCATCATGATTTTCTCTAATATTTACTTCTTACCTGTTTGTGCAAGAACCA---AAACCA 4332  
DB 4912 GTGTTACTCATCTTTATTTGTTCTCCATGATTTATCTAGCTTGCAAAAAGCCAAAGAACCC 4971  
QY 4333 CATCAAGCACAACCTCCCAACAGAGCCCTTGACCTTAGCTTACGATGAGACTTAGAC 4392  
DB 4972 CAAGAAGGCAAAACAACCGACTGAACCTTTAACTTAGCTATGATGAGATGCCGAC 5031  
QY 4393 ATGTAA 4398  
DB 5032 ATGTAA 5037  
RESULT 2  
US-09-919-172-96  
; Sequence 96, Application US/09919172  
; Patent No. 6673545  
; GENERAL INFORMATION:  
; APPLICANT: Paris, Mary  
; APPLICANT: Turner, Christopher M.  
; TITLE OF INVENTION: PROSTATE CANCER MARKERS  
; FILE REFERENCE: PA-0036 US  
; CURRENT APPLICATION NUMBER: US/09/919,172  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: 60/222,469  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PERL Program  
; SEQ ID NO 96  
; LENGTH: 579  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. 6673545 041764.1  
US-09-919-172-96

Query Match 3.7%; Score 161.8; DB 4; Length 579;  
Best Local Similarity 58.7%; Pred. No. 1.4e-42;  
Matches 340; Conservative 0; Mismatches 227; Indels 12; Gaps 3;  
QY 2662 GAAAAACCTTACAATGGAGGACGACCATGTCCCAACTGGATCTCAAGATCAGGTAC-- 2719  
DB 1 GAAAAACCATATAATGGAGGAAGGCGCTTGCCCAAACTGGACCATGTCAACAGGACAG 60  
QY 2720 ----ATGGGAGTCCCATGTTACAGTGAAGTCAATCAGTATTCTCGTGTGTAGACAC 2775  
DB 61 GTGTATGAGGTGTCCCATGCCACAGTCACTGCAACCAAGTACCTATGGGTCAAGAGCCC 120  
QY 2776 TGGTCTTCATCAAAATCAACAATGAGCTGAGGTCCCTGCGCTGTGGAGGAGAACACAA 2835  
DB 121 TGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAGAACTGTGGAGAGGGGTCAA 180  
QY 2836 TCTAGGAAATCAGATGTGT---GAATCTGGGATGTGTGAAGTGGAGCGAGTGGATAGC 2892  
DB 181 ACCGAAAAGTGAGATGATGTCAGAAATACAGACATGCGCCCTTCTGAACATGTAGAGAT 240  
QY 2893 AACCTGTGCAACGAGATGAATTCGCCAGAAACCCAGTCTGTCTCTTATGTGTCCTC 2952  
DB 241 TACCTCTGTGACCCAGAGAGATGTCCTCTGGGCTTAGAGTGTGCAATPACCATGCGCT 300  
QY 2953 AATGAGTGTGTATGTCTGAGTGGGAGATTTGGAGCAAAATGCCACAGTCAATGCGATCCC 3012  
DB 301 GAGGACTGTGTATATCTGAATGGGCTCCATGGACCCCAATGTGTTTTCCTTGAATCAA 360  
QY 3013 CACACAAT---CGAGAGAGACTGCCACCTGCTTAAGCAACCATCACTGAATCAAGGACT 3069  
DB 361 AGCAGTTTCCGCAAGGTGACTGATCCCATCAAGCAACCAAGTGTGAAGGAAGATCT 420  
QY 3070 TGTGCTGAAGACTCAAGGTGACGCTTGCCTCTCTGAATGAAATTTGCTTCCAGTTCCAG 3129  
DB 421 TGGCCTTAATGCTGTTGAGAAAGAACCCCTGTACCTGACAAACAACTGTACCACATGAT 480

QY 3130 TACAATCTAACAGAGTGGAGCACATGCGAGTGAAGTGAAGAAACGACCCCTGTGTCAGGC 3189  
DB 481 TATAATGTAAACAGACTGGAGTACATGTGAGTGAAGAGGAGGAGTTTGTGGAATGGA 540  
QY 3190 GTCAGGAGCCGCTCTAAGCTGTGCTGTGTCAGTGTATGCG 3228  
DB 541 ATAAAAACAAGATGTTGATGTTGTTGATGATGATGCG 579  
RESULT 3  
US-07-862-021B-11  
; Sequence 11, Application US/07862021B  
; Patent No. 5279866  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M  
; APPLICANT: Klar, Avihu  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham  
; STREET: 30 Rockefeller Plaza  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/862,021B  
; FILING DATE: 19920405  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 40028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 664-0525  
; TELEX: 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3226 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 136...2543  
US-07-862-021B-11

Query Match 1.0%; Score 44; DB 1; Length 3226;  
Best Local Similarity 58.3%; Pred. No. 0.0047;  
Matches 77; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 2566 TGCCCATTTGATTGCAAGTTAAGCGATTGCTAGTTGGGGTCTTTCAGTTTCATCTTGT 2625  
DB 2113 TGCCCCATTTGATGTGAGCTAACGGAGTGTGCCAGTGTGTCGAGTGCATACCTCTGT 2172  
QY 2626 GGAATTTGGAGTGAAGATTCGATCCAAATGGCTTAAAGAAAAACCTTACATGGAGACA 2685  
DB 2173 GGGAGGGCCACATGATCAGAAACAAGATGATCAAAATAGAACCAACAGTTTGGAGACA 2232  
QY 2686 CCATGTCCCAAA 2697  
DB 2233 GCATGCCCAAA 2244

RESULT 4

IS-08-313-288B-11  
Sequence 11, Application US/08313288B  
Patent No. 5750502

GENERAL INFORMATION:  
APPLICANT: Jessell, Thomas M. and Avihu Klar  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/313,288B  
FILING DATE: January 5, 1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0526  
TELEX:

INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3226 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 136..2543  
PCT-US93-03164-11

Query Match 1.0%; Score 44; DB 1; Length 3226;  
Best Local Similarity 58.3%; Pred. No. 0.0047;  
Matches 77; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 2566 TGCCCATTTGCAAGTTAAGCGATTGCTAGTTGGGGGCTTTCAGTTCACTTGT 2625  
DB 2113 TGCCCATTTGCTGAGCTAAGCGAGTGCTCCAGTGTCGAGTGCAATACCTCTGT 2172

QY 2626 GGAATTGGAGTGAGATTCGATCCAAATGCTAAAGAAAACCTTACAATGGAGGACGA 2685  
DB 2173 GGGAGGGCCACATGATCAGAACAAAGATGATCAAAATAGAACACACAGTTTGGAGGAACA 2232

QY 2686 CCATGTCCTCCAAA 2697  
DB 2233 GCATGCCCAGNA 2244

## RESULT 5

PCT-US93-03164-11  
Sequence 11, Application PC/TUS9303164  
GENERAL INFORMATION:  
APPLICANT: Jessell, Thomas M  
APPLICANT: Klar, Avihu  
TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 30 Rockefeller Plaza

CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03164  
FILING DATE: 19930402  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3226 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:

NAME/KEY: CDS  
LOCATION: 136..2543  
PCT-US93-03164-11

Query Match 1.0%; Score 44; DB 5; Length 3226;  
Best Local Similarity 58.3%; Pred. No. 0.0047;  
Matches 77; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 2566 TGCCCATTTGCTGCAAGTTAAGCGATTGCTAGTTGGGGGCTTTCAGTTCACTTGT 2625  
DB 2113 TGCCCATTTGCTGAGCTAAGCGAGTGCTCCAGTGTCGAGTGCAATACCTCTGT 2172

QY 2626 GGAATTGGAGTGAGATTCGATCCAAATGCTAAAGAAAACCTTACAATGGAGGACGA 2685  
DB 2173 GGGAGGGCCACATGATCAGAACAAAGATGATCAAAATAGAACACACAGTTTGGAGGAACA 2232

QY 2686 CCATGTCCTCCAAA 2697  
DB 2233 GCATGCCCAGAA 2244

## RESULT 6

US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFELINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION NUMBER: US/08/232.463  
FILING DATE: 26-AUG-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935.313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 1.0%; Score 42.8; DB 1; Length 7218;  
Best Local Similarity 4.3%; Pred. No. 0.022;  
Matches 17; Conservative 210; Mismatches 167; Indels 0; Gaps 0;

QY 1780 AAAAAGACTGATTTGACCTTTTACGATGAGTGGACACCTGCCCAAGGATGCGCAA 1859  
DB 1426 RR 1367  
QY 1840 GCAGGAATGCCAGTAAACAGTCTCGATACAGATCATCAAGAGAGAGCCAAAT 1899  
DB 1366 RR 1307  
QY 1900 GGAGCCAGGAATGCCAGTAAACAGTCTCGATACAGATCATCAAGAGAGAGTCT 1959  
DB 1306 RR 1247  
QY 1960 TGTCTGTATATCGGTGGAAGCAGACAGAAATGGAGCCCTTCATCTTAGTCCCAAGTCT 2019  
DB 1246 RR 1197  
QY 2020 GTCTGCGAGGAATACGGGACAGTGAAGCTTGGAAGGGTTACAAACAGAGCT 2079  
DB 1186 RR 1127  
QY 2080 GTCTCATGATCTCTGATGACACACCGGTGACAGAAATGATGATGCTCAAGAGACA 2139  
DB 1126 RR 1067  
QY 2140 AACGGATGCTCTCTCTGTGCAAGATGCACAG 2173  
DB 1066 ATCGAAGTCTCTGACCTGACCTGACGACCAAGTCTGG 1033

## RESULT 7

US-09-132-769-4  
Sequence 4, Application US/091327694  
Patent No. 6525023  
GENERAL INFORMATION:  
APPLICANT: Motoco Yamasaki  
APPLICANT: Kenji Shibata  
APPLICANT: No. 6525023uo Hanai  
APPLICANT: Akiko Furuya  
APPLICANT: Kaoru Miyamoto  
TITLE OF INVENTION: NOVEL VASCULAR SMOOTH MUSCLE CELL GROWTH FACTOR

FILE REFERENCE: 11078  
CURRENT APPLICATION NUMBER: US/09/132,769A  
CURRENT FILING DATE: 1998-08-12  
EARLIER APPLICATION NUMBER: HEI9-218491  
EARLIER FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4  
LENGTH: 4061  
TYPE: DNA  
ORGANISM: BOVINE  
US-09-132-769-4

Query Match 0.9%; Score 39.8; DB 4; Length 4061;

Best Local Similarity 56.5%; Pred. No. 0.14; 57; Indels 0; Gaps 0;  
Matches 74; Conservative 0; Mismatches 0;

QY 103 TGTCCCGGGATTGTGTAGTATCTGAGTTCTTACCATGGTCCAACTGTAGCAAGGATGT 162  
DB 2119 TGCCCCATTGACTGTGAGCTCAGCGAGTGGTCCAGTGGTGGAAATGTAAACAAGTCATGT 2178  
QY 163 GGAAGAAATTCAGCATAGAACTCGCGGGTCATAGTCCCCCTCTCTTTGGTGGTTG 222  
DB 2179 GGAAGAGGACATCATGATTCGAACCCCGCATGATCCAAATGGAGCCTCAGTTTGGAGGCACA 2238  
QY 223 CAATGTCCAAA 233  
DB 2239 CCTGCCCCAGA 2249

## RESULT 8

US-09-132-769-2  
Sequence 2, Application US/09132769A  
Patent No. 6525023  
GENERAL INFORMATION:  
APPLICANT: Motoco Yamasaki  
APPLICANT: Kenji Shibata  
APPLICANT: No. 6525023uo Hanai  
APPLICANT: Akiko Furuya  
APPLICANT: Kaoru Miyamoto  
TITLE OF INVENTION: NOVEL VASCULAR SMOOTH MUSCLE CELL GROWTH FACTOR  
FILE REFERENCE: 11078  
CURRENT APPLICATION NUMBER: US/09/132,769A  
CURRENT FILING DATE: 1998-08-12  
EARLIER APPLICATION NUMBER: HEI9-218491  
EARLIER FILING DATE: 1997-08-13  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 2  
LENGTH: 3079  
TYPE: DNA  
ORGANISM: human  
US-09-132-769-2

Query Match 0.9%; Score 39.2; DB 4; Length 3079;

Best Local Similarity 50.5%; Pred. No. 0.17; 93; Indels 0; Gaps 0;  
Matches 95; Conservative 0; Mismatches 0;

QY 46 GAAATATGCAACACTTTGCCCTTCAGCCTCCTACAGAACAGGCTTGCTCATCTTGT 105  
DB 1960 GAACTTGGAGACTGCAATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCCCTGAATGC 2019  
QY 106 CCCCCGGATTGTGTAGTATCTGAGTTCTTACCATGGTCCAACTGTAGCAAGGATGTGGG 165  
DB 2020 CCATTTGACTGTGAGCTCACCGAGTGGTCCCAAGTGGTGGAAATGTAAACAAGTCATGTGGG 2079  
QY 166 AAGAAATTCAGCATAGAACTCGCGGGTCATAGTCCCCCTCTCTTTGGTGGTTGCAA 225  
DB 2080 AAGGCCAGTGTATTCGAACCCCGCATGATCCAAATGGAGCCTCAGTTTGGAGGTGCAACC 2139  
QY 226 TGTCCTAAA 233  
DB 2140 TGCCCCAGA 2147

```

; RESULT 9
; US-09-640-173-184
; Sequence 184, Application US/09640173
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 184
; LENGTH: 3079
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-640-173-184

Query Match      0.9%; Score 39.2; DB 4; Length 3079;
Best Local Similarity 50.5%; Pred. No. 0.17; 93; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 0;

2Y 46 GAAATATGCGAACACTTGGCCCTTCAGCTTACAGAACAGGCTTGCCTCATTCCTTGT 105
Db 1960 GAACCTTGGAGACTGCAATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCCCTGAATGC 2019
2Y 106 CCCGGGATGTGTAGTATCTGAGTTCTTACCATGCTCCAACTGTCCAGTACGAGGATGTGG 165
Db 2020 CCATTGACTGTGAGCTCACCGAGTGGTCCCACTGTGCGAATGTAAAGATCATGTGGG 2079
2Y 166 AAGAAATTGCGACATAGAACTCCGCGGGTCATAGCTCCCCCTCTCTTTGGTGGTTGCAA 225
Db 2080 AAGGCCACGTGATTTCGAACCCGGATGATCCAAATGGAGCCTCAGTTTGAGGTGCACCC 2139
2Y 226 TGTCCAAA 233
Db 2140 TGCCCCAGA 2147

; RESULT 10
; US-09-713-550-184
; Sequence 184, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 184
; LENGTH: 3079
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-713-550-184

Query Match      0.9%; Score 39.2; DB 4; Length 3079;
Best Local Similarity 50.5%; Pred. No. 0.17; 93; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 0;

2Y 46 GAAATATGCGAACACTTGGCCCTTCAGCTTACAGAACAGGCTTGCCTCATTCCTTGT 105
Db 1960 GAACCTTGGAGACTGCAATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCCCTGAATGC 2019
2Y 106 CCCGGGATGTGTAGTATCTGAGTTCTTACCATGCTCCAACTGTCCAGTACGAGGATGTGG 165
Db 2020 CCATTGACTGTGAGCTCACCGAGTGGTCCCACTGTGCGAATGTAAAGATCATGTGGG 2079
2Y 166 AAGAAATTGCGACATAGAACTCCGCGGGTCATAGCTCCCCCTCTCTTTGGTGGTTGCAA 225
Db 2080 AAGGCCACGTGATTTCGAACCCGGATGATCCAAATGGAGCCTCAGTTTGAGGTGCACCC 2139
2Y 226 TGTCCAAA 233
Db 2140 TGCCCCAGA 2147

; RESULT 11
; US-09-640-173-177
; Sequence 177, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640,173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 177
; LENGTH: 3999
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-640-173-177

Query Match      0.9%; Score 39.2; DB 4; Length 3999;
Best Local Similarity 50.5%; Pred. No. 0.21; 93; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 0;

2Y 46 GAAATATGCGAACACTTGGCCCTTCAGCTTACAGAACAGGCTTGCCTCATTCCTTGT 105
Db 1387 GAACCTTGGAGACTGCAATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCCCTGAATGC 1446
2Y 106 CCCGGGATGTGTAGTATCTGAGTTCTTACCATGCTCCAACTGTCCAGTACGAGGATGTGG 165
Db 1447 CCATTGACTGTGAGCTCACCGAGTGGTCCCACTGTGCGAATGTAAAGATCATGTGGG 1506
2Y 166 AAGAAATTGCGACATAGAACTCCGCGGGTCATAGCTCCCCCTCTCTTTGGTGGTTGCAA 225
Db 1507 AAGGCCACGTGATTTCGAACCCGGATGATCCAAATGGAGCCTCAGTTTGAGGTGCACCC 1566
2Y 226 TGTCCAAA 233
Db 1567 TGCCCCAGA 1574

; RESULT 12
; US-09-713-550-177
; Sequence 177, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 177
; LENGTH: 3999
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-713-550-177

Query Match      0.9%; Score 39.2; DB 4; Length 3999;
```

Best Local Similarity 50.5%; Pred. No. 0.21;  
Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
QY 46 GAAATATCGAACAACCTTTGCCCTTCAGCCTCTCTACAGAACAGGCTTGCCCTCATTCCTTGT 105  
DB 1387 GAACTTGGAGAGCTGCAATGAGGATCTGGAGCAGGTGGAGAGAGTGCATGCTCCTCGAATGC 1446  
QY 106 CCCCGGATTGTTAGTATCTGAGTCTTACCATGTCCTCAACTGTAGCAAGGATGTGGG 165  
DB 1447 CCCATTGACTGTGAGCTCACCGAGTGTCCAGTGTGCGGAATGTACAAAGTCAATGTGGG 1506  
QY 166 AAGAAATTCAGCATAGAACTGCGCGGCTCATAGTCTCCCTCTCTCTTTGGTGTTCGAA 225  
DB 1507 AAAGCCACGCTGATTCGAACCCGGATGATCCAAATGGAGCCTCAGTTTGGAGGTGCACCC 1566  
QY 226 TGTCCTCAA 233  
DB 1567 TGCCCA 1574

RESULT 13  
US-09-976-594-353  
; Sequence 353, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 353  
; LENGTH: 4866  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 415378.3  
US-09-976-594-353

Query Match 0.9%; Score 39.2; DB 4; Length 4866;  
Best Local Similarity 50.5%; Pred. No. 0.25; Indels 0; Gaps 0;  
Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
QY 46 GAAATATCGAACAACCTTTGCCCTTCAGCCTCTCTACAGAACAGGCTTGCCCTCATTCCTTGT 105  
DB 1964 GAACTTGGAGAGCTGCAATGAGGATCTGGAGCAGGTGGAGAGTGCATGCTCCTGAATGC 2023  
QY 106 CCCCGGATTGTTAGTATCTGAGTCTTACCATGTCCTCAACTGTAGCAAGGATGTGGG 165  
DB 2024 CCATTTGACTGTGAGCTCCAGGAGTGTCCCGAGTGTCCCGAGTGTGGAATGTAAAGTCAATGTGGG 2083  
QY 166 AAGAAATTCAGCATAGAACTGCGCGGCTCATAGTCTCCCTCTCTCTTTGGTGTTCGAA 225  
DB 2084 AAGGCCACGCTGATTCGAACCCGGATGATCCAAATGGAGCCTCAGTTTGGAGGTGCACCC 2143  
QY 226 TGTCCTCAA 233  
DB 2144 TGCCCA 2151

RESULT 14  
US-09-621-976-15639/c  
; Sequence 15639, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 15639  
; LENGTH: 505  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-621-976-15639  
Query Match 0.8%; Score 37; DB 4; Length 505;  
Best Local Similarity 14.2%; Pred. No. 0.22;  
Matches 55; Conservative 160; Mismatches 170; Indels 2; Gaps 1;  
QY 1089 TCAGGGGAAAAAGGATTTAGAACGAGGAGGCGCATGTCCTCATGGAATCTACAGGGCC 1148  
DB 443 TTAACRWAWAGAGAAWKKWAGGWRACARAGTTWAGACWAMARRGRCARGTGS-- 386  
QY 1149 TGAGGGGCAATGCCCTCATTTTGGTGGAGTCTCTTGTGAGGATCCAAATGTGTACCG 1208  
DB 385 WGSWKGGYRWGKGMMAAAKRRMAAAGGSGYCGMTSYTSGKMTGKSGMTKIKRMWTYSG 326  
QY 1209 ATGGCTGGCATCAGAGGATCTGTTCCCTCATCATGGAATGTGGCTGGGACATCG 1268  
DB 325 MMWTSYKTK 266  
QY 1269 TATCTGGAAGGCGCTGCGCAAGTACGCGGAGAGATGTATCAGGAGTCTTTGGCC 1328  
DB 265 MMSASAYRARRSMYGARRSMRAGAGWARRGKRRGKRRGKRRGKRRGKRRGKRRGKRRG 206  
QY 1329 AGTTCCCTCTCTCTGAGAGAGTCTTGTGAATTCCTCCCGAATGAGTGTGTCT 1388  
DB 205 CREWMWSCRMYSYSCMSGCRGTCAKMWRYAKRYASSMKYMMGCRKWCYAKCAR 146  
QY 1389 GAGCGAGTGCAGGAGTGTCT 1448  
DB 145 MYGYERSRSSTGSRGMYRRKMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMYMY 86  
QY 1449 GAAACAGACAGTCAAGAACTATCTCT 1475  
DB 85 CKMCSRMMKMSMMMSMMMSMMMSMMMSMMMSMMMSMMMSMMMSMMMSMMMSMMMSMM 59

RESULT 15  
US-09-397-992A-25  
; Sequence 25, Application US/09397992A  
; Patent No. 6329175  
; GENERAL INFORMATION:  
; APPLICANT: Conklin, Darrell  
; APPLICANT: Grant, Francis J.  
; APPLICANT: Rixon, Mark W.  
; APPLICANT: Kindvogel, Wayne  
; TITLE OF INVENTION: Interferon-epsilon  
; FILE REFERENCE: 98-46  
; CURRENT APPLICATION NUMBER: US/09/397,992A  
; CURRENT FILING DATE: 1999-09-16  
; PRIOR APPLICATION NUMBER: 60/101,012  
; PRIOR FILING DATE: 1998-09-18  
; PRIOR APPLICATION NUMBER: 60/118,578  
; PRIOR FILING DATE: 1999-02-05  
; PRIOR APPLICATION NUMBER: 60/142,766  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 25  
; LENGTH: 576  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: This degenerate sequence encodes the amino acid  
; OTHER INFORMATION: sequence of SEQ ID NO:24.

NAME/KEY: variation  
LOCATION: (1)...(576)  
OTHER INFORMATION: N is any nucleotide.  
IS-09-397-992A-25

Query Match 0.8%; Score 37; DB 4; Length 576;  
Best Local Similarity 30.0%; Pred. No. 0.25;  
Matches 65; Conservative 35; Mismatches 117; Indels 0; Gaps 0;

Y	3988	CGGCGTGTGATCCAGCCTGCAGAAACCTTCTCTACTGTACACAGGGTGGACTCTGT	4047
b	233	CNGTGTNCAYGARATHYTNCAFCARATHYTYACNYTNTNCARACNCAYGNCNATGG	292
Y	4048	GGTGTGAGAAGGGCTATACAGAGATAATGAAATCAAATGGTTTCTGTGATTTACTGCATG	4107
b	293	GNATHGGGARGARAAYCAVATHGAPARGTNYTNGCNGCNVTNCAYMGCNARYTNGART	352
Y	4108	AAAGTACCAGGCTCAGAGGATAAAGAGCTGATGTGAAAACCTTCTGGGAAAAACAGA	4167
b	353	AYGTNGARWSNTYNGGNGNYTNAAYGCNGCNCARAARWSNGNGNWSNWSNWCNCARA	412
Y	4168	CCTGTCAATTCAAAAATACATGATATTTTAAAGGAT	4204
b	413	AYTNGNYTNCARATHAAGCNTAYTTYMGNAT	449

Search completed: February 23, 2004, 20:35:45  
Job time : 336.257 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1232.8	28.0	4810	15	US-10-074-566-40	Sequence 40, Appl
2	1232.8	28.0	5669	15	US-10-074-566-59	Sequence 59, Appl
3	1232.8	28.0	6373	13	US-10-000-512-1	Sequence 1, Appl
4	1232.8	28.0	6373	15	US-10-074-566-1	Sequence 58, Appl
5	1232.8	28.0	6373	15	US-10-074-566-58	Sequence 332, Appl
6	1229.6	28.0	6378	14	US-10-037-270-332	Sequence 332, Appl
7	1229.6	28.0	6378	15	US-10-117-722-332	Sequence 958, Appl
8	580.6	13.2	3053	15	US-10-104-047-958	Sequence 20, Appl
9	512.4	11.7	516	9	US-09-960-253-20	Sequence 27172, A
10	460	10.5	744	9	US-09-864-761-27172	Sequence 5068, Ap
C 11	239.2	5.4	530	14	US-10-029-386-5068	Sequence 1824, A
C 12	235	5.3	235	14	US-10-029-386-1824	Sequence 8718, Ap
C 13	201	4.6	583	14	US-10-029-386-8718	Sequence 22453, A
C 14	174	4.0	174	14	US-10-029-386-22453	Sequence 96, Appl
15	161.8	3.7	579	9	US-09-919-172-96	

SEQ ID NO 40  
LENGTH: 4810  
TYPE: DNA  
ORGANISM: human  
US-10-074-566-40

Query Match  
Best Local Similarity 57.7%; Pred. No. 0;  
Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;

10 TGCATTCAGAGCTGAACCGAAGCTGTGGTGGCAATGAATATATGCGAAGCTTTGCCCTT 69  
Db TGCATTCAGAGCTGAACCGAAGCTGTGGTGGCAATGAATATATGCGAAGCTTTGCCCTT 69

70 CAGCTCTCTACAGACAGCTTGCCTCAATCTTGTCCCGGGATTTGTAGTATCTGAG 129  
Db CAGCTCTCTACAGACAGCTTGCCTCAATCTTGTCCCGGGATTTGTAGTATCTGAG 129

348 AAGCTCTCTGGAGCAGCTTGCCTCAATCTTGTCCCGAGAGATTTGATGCTGTGAA 407  
Db AAGCTCTCTGGAGCAGCTTGCCTCAATCTTGTCCCGAGAGATTTGATGCTGTGAA 407

130 TTCTTACCATGCTCAACTGTAGCAAGGATGTGGGAAGAAATTCAGCATAGAACTCGC 189  
Db TTCTTACCATGCTCAACTGTAGCAAGGATGTGGGAAGAAATTCAGCATAGAACTCGC 189

408 TTTTCTGCTGTGTCGGAATGCTCCAAAGACCTGGCGACGGGCTCCAGCACCGGACGGT 467  
Db TTTTCTGCTGTGTCGGAATGCTCCAAAGACCTGGCGACGGGCTCCAGCACCGGACGGT 467

190 GGGTCATAGCTCCCTCTCTTTGGTGTGTGCAATGTCCAAATCTGACTGAGTCAAGA 249  
Db GGGTCATAGCTCCCTCTCTTTGGTGTGTGCAATGTCCAAATCTGACTGAGTCAAGA 249

468 CATGTGGTGGCCCCCGAGTTGGAGGCTCTGGCTGTCCAAACCTGACGAGTTCCAG 527  
Db CATGTGGTGGCCCCCGAGTTGGAGGCTCTGGCTGTCCAAACCTGACGAGTTCCAG 527

250 GCTGTGATGCTCCCAATTTCTGCTCTTGGGGAAGAGGAATATACATTTAGCCTTAA 309  
Db GCTGTGATGCTCCCAATTTCTGCTCTTGGGGAAGAGGAATATACATTTAGCCTTAA 309

528 GTGTG-----CCAAATCCAGTCCATGCGAGGCGGAGGAGCTCAGGTACAGCTGCAT 578  
Db GTGTG-----CCAAATCCAGTCCATGCGAGGCGGAGGAGCTCAGGTACAGCTGCAT 578

310 GTTGAACATGAGCTAAATGAGACTGCTCATCTTAAGAAATTAATCCAGCGGAGA 369  
Db GTTGAACATGAGCTAAATGAGACTGCTCATCTTAAGAAATTAATCCAGCGGAGA 369

579 GTGGGCGCTGAGACCTGCTCAATGCCCCACTCCGACAGTAAGACAAGCAAGGAGA 638  
Db GTGGGCGCTGAGACCTGCTCAATGCCCCACTCCGACAGTAAGACAAGCAAGGAGA 638

370 ACTGTTCTGGATTTAA-----CTCTGATTTCAAATGAG 402  
Db ACTGTTCTGGATTTAA-----CTCTGATTTCAAATGAG 402

639 CCGGGGAAGATAAAGAACGGGAAGACCGCACCAAGAGTAAGGATCCAGAGGCC 698  
Db CCGGGGAAGATAAAGAACGGGAAGACCGCACCAAGAGTAAGGATCCAGAGGCC 698

403 CGAGTCACCTTTAAACATCAAAAGTTACA-----AAGCACATCATCTCGAAGTCT 453  
Db CGAGTCACCTTTAAACATCAAAAGTTACA-----AAGCACATCATCTCGAAGTCT 453

699 CCGAGCTTTAATAAGAAAAGAGAACAACAGCAAGCAGAGAGAGTATGCAATTAACAGACGAGAA 758  
Db CCGAGCTTTAATAAGAAAAGAGAACAACAGCAAGCAGAGAGAGTATGCAATTAACAGACGAGAA 758

454 TGGGCAATAGACATAGGTATCAACCCGCGAGGTTTCGTGTACAAGAGTGTATGCAAA 513  
Db TGGGCAATAGACATAGGTATCAACCCGCGAGGTTTCGTGTACAAGAGTGTATGCAAA 513

759 TGGACATCCAGATTGGAATACAGACAGAGAGTTATGCAATTAACAGACGAGAA 818  
Db TGGACATCCAGATTGGAATACAGACAGAGAGTTATGCAATTAACAGACGAGAA 818

514 AATGTATGTTAAGCTTTGCCCTTCAAGATTCTTCCCATGACTGTTTCACTGCAATC 573  
Db AATGTATGTTAAGCTTTGCCCTTCAAGATTCTTCCCATGACTGTTTCACTGCAATC 573

819 GCTGTGATTTAAGCTTTGCCAGCAAGAGAGCTTCCAATGACCTTCCAGTCTCTGTG 878  
Db GCTGTGATTTAAGCTTTGCCAGCAAGAGAGCTTCCAATGACCTTCCAGTCTCTGTG 878

574 ATGCCCAAGCTGTGAACCTCCAGTGTCTCTCTGGGCGGCTGCTCCAGACATGC 633  
Db ATGCCCAAGCTGTGAACCTCCAGTGTCTCTCTGGGCGGCTGCTCCAGACATGC 633

879 ATCAACAAAGATGCGAGGTTCCGAGTGTGTCAGAGTGGAGCCCTGCTCAAAACATGC 938  
Db ATCAACAAAGATGCGAGGTTCCGAGTGTGTCAGAGTGGAGCCCTGCTCAAAACATGC 938

634 CGTTTCAAGGAGTCTCTTGCAGGATTTAGGAGCAGGAGCGGAAAGTGAAGCATGCT 693  
Db CGTTTCAAGGAGTCTCTTGCAGGATTTAGGAGCAGGAGCGGAAAGTGAAGCATGCT 693

939 CATGACATGTTGCCCTCAGGCACTGTGTAAAGACACGAACCATCAGCAGTTTCCC 998  
Db CATGACATGTTGCCCTCAGGCACTGTGTAAAGACACGAACCATCAGCAGTTTCCC 998

694 ATTGGAGTGGAAAGAGTGTCTGAACTTTCTTGAAGAGAGGCTGCAATTTGTAAGGA 753  
Db ATTGGAGTGGAAAGAGTGTCTGAACTTTCTTGAAGAGAGGCTGCAATTTGTAAGGA 753

999 ATTGGCAGTGAAGAGGAGTGTCCAGAAATTTGAAGAAAAGAACCTGTTGTCTCAAGGA 1058  
Db ATTGGCAGTGAAGAGGAGTGTCCAGAAATTTGAAGAAAAGAACCTGTTGTCTCAAGGA 1058

754 GAACTTCTCAGCAATGTCCTCAGGATTTCTCTGGAGAACTTCTGAAATGAAGAAATGCCAA 813  
Db GAACTTCTCAGCAATGTCCTCAGGATTTCTCTGGAGAACTTCTGAAATGAAGAAATGCCAA 813

1059 GATGGAGTTGTCCTCTGTGCGCAGTATGGCTGGAGAACTACAGAGTGAAGTGGCTG 1118  
Db GATGGAGTTGTCCTCTGTGCGCAGTATGGCTGGAGAACTACAGAGTGAAGTGGCTG 1118

814 GTCTCTCTCTCTCAGCAGAGATCCCCACTGGCATGTGACGGGACCGCTGTGTGGC 873  
Db GTCTCTCTCTCTCAGCAGAGATCCCCACTGGCATGTGACGGGACCGCTGTGTGGC 873

1119 GTGGAACCTTTGTCTAGTCAGCAGGACAGAGGCGCGCAACACAGACGCGCTCTCTGTGGA 1178  
Db GTGGAACCTTTGTCTAGTCAGCAGGACAGAGGCGCGCAACACAGACGCGCTCTCTGTGGA 1178

874 GTGGGATCCAGACCGGAGTGTACTGTGCCAG-----AGGTACACCA 921  
Db GTGGGATCCAGACCGGAGTGTACTGTGCCAG-----AGGTACACCA 921

1179 GGGGGCATCCAGACCCGAGAGTGTACTGCTGGAGGCCAACGAAACCTCTCTCACAA 1238  
Db GGGGGCATCCAGACCCGAGAGTGTACTGCTGGAGGCCAACGAAACCTCTCTCACAA 1238

922 GCTGCGCAGCTGAGGCGCAAGGAGCTCTCTAGACCTGTGGAAAAGGCATTATGTGGGA 981  
Db TTAAGTACCCACAGAAACAAAGAGCCCTCAAGCCAAATGGACTTAAATTTATGACTGGA 1298

982 CCGCGCCGCTTCCCTCTCAGCTCTGCAATATCCCTGCTCTACGAGCTGCATATCTCT 1041  
Db CTTATCCCTTAATACACAGCTGTGCCACATTCCTTGTCCAACTGAATGTGAAGTTTCA 1358

1042 TCTGTGACGCTGGGCTGTGCATCCATGAATCTGATGAACCTGATGAACCTCAGGGGAAAAA 1101  
Db TCTGTGACGCTGGGCTGTGCATCCATGAATCTGATGAACCTGATGAACCTCAGGGGAAAAA 1101

1359 CTTTGTGAGCTTGGGACCTTGTACTTATGAATACTGTATATGATCAGCAGGAGAAAAA 1418  
Db CTTTGTGAGCTTGGGACCTTGTACTTATGAATACTGTATATGATCAGCAGGAGAAAAA 1418

1102 GGAATTTAGAACAGGACGCG-----CCATGTCTCTCATGGAATCTACAGGGCTGCA 1152  
Db GGAATTTAGAACAGGACGCG-----CCATGTCTCTCATGGAATCTACAGGGCTGCA 1152

1419 GGTCTCAAACTCAGGAAGCGGCGCATTAACCAATGAGCCACTGGAGGCTCTGGGTAACC 1478  
Db GGTCTCAAACTCAGGAAGCGGCGCATTAACCAATGAGCCACTGGAGGCTCTGGGTAACC 1478

1153 GGGCATTTGCCCTCANTTTGGTGAAGTCTGTTCTCTGAGGATCCCAATGTGCTACCGATGG 1212  
Db GGGCATTTGCCCTCANTTTGGTGAAGTCTGTTCTCTGAGGATCCCAATGTGCTACCGATGG 1212

1479 GGAATCTGCCCTCCTACTACTGGAAGCCATTCCTCTGGAAGGCTGCTGTTTATGACTGG 1538  
Db GGAATCTGCCCTCCTACTACTGGAAGCCATTCCTCTGGAAGGCTGCTGTTTATGACTGG 1538

1213 ----CTGECATCAGAAAGGATCTGTTTCCCTGATCATGGAA----AATGTGGCCTGGGACAT 1266  
Db ----CTGECATCAGAAAGGATCTGTTTCCCTGATCATGGAA----AATGTGGCCTGGGACAT 1266

1539 AAGCGTGAAGCTGGGACCTGCGAGCCAGATAACGGAAGAGAGTGTGTCAGGACAG 1598  
Db AAGCGTGAAGCTGGGACCTGCGAGCCAGATAACGGAAGAGAGTGTGTCAGGACAG 1598

1267 CGTATTTGAAAGCGCTGTGCGAGAAATGACCGGAGAGAGATGTATCAGGAGTCTTTGC 1326  
Db CGTATTTGAAAGCGCTGTGCGAGAAATGACCGGAGAGAGATGTATCAGGAGTCTTTGC 1326

1599 CAAGTTCAAGAGGTTGTGTGCATCAACAGTGTGAGAGAGAGTTGACAGACAGCTGTGC 1658  
Db CAAGTTCAAGAGGTTGTGTGCATCAACAGTGTGAGAGAGAGTTGACAGACAGCTGTGC 1658

1327 CCAAGTTCCCTCTCTGAGAGGAGTCTTGTGAAATTCCTGCGCAATGAGTGTGTG 1386  
Db CCAAGTTCCCTCTCTGAGAGGAGTCTTGTGAAATTCCTGCGCAATGAGTGTGTG 1386

1659 AGAGATGCCATCTTCCCCATCTCTGTGGCTGTGATGCCCATGCCGAAAGACTGTGTG 1718  
Db AGAGATGCCATCTTCCCCATCTCTGTGGCTGTGATGCCCATGCCGAAAGACTGTGTG 1718

1387 CTGAGCGATGGAAGGAGTGTGATCTCTGTTCCAGTCTCTGTTTCAATAAAGTCTGAT 1446  
Db CTGAGCGATGGAAGGAGTGTGATCTCTGTTCCAGTCTCTGTTTCAATAAAGTCTGAT 1446

1719 CTGAGCAGTGTGTCTGAGTGTCTCTCTGCTCACACACCTGCTCAGGGAACAGCAGAA 1778  
Db CTGAGCAGTGTGTCTGAGTGTCTCTCTGCTCACACACCTGCTCAGGGAACAGCAGAA 1778

1447 GGGAAACAGACAGGCTCAAGAACTATCTCTGCG-----ACTGCTGGGGAAGCTGGAAGCCA 1503  
Db GGGAAACAGACAGGCTCAAGAACTATCTCTGCG-----ACTGCTGGGGAAGCTGGAAGCCA 1503

1779 GGGAAACAGATAGGACAGATCTCTCTGCGCTATGCGGTGAGAGAGTGGAAATTCG 1838  
Db GGGAAACAGATAGGACAGATCTCTCTGCGCTATGCGGTGAGAGAGTGGAAATTCG 1838

1504 TGTCCCCCTAGTCAGGCTCTCCAAGAGCATCGTTTGTGTAATGACCATTCCTGTATGCA 1563  
Db TGTCCCCCTAGTCAGGCTCTCCAAGAGCATCGTTTGTGTAATGACCATTCCTGTATGCA 1563

1839 TGTCCAAATAGCAGTCTTGTCAAGAAAGTACGAAGCTGTAAATGAGCATCTTGTGACAGTG 1898  
Db TGTCCAAATAGCAGTCTTGTCAAGAAAGTACGAAGCTGTAAATGAGCATCTTGTGACAGTG 1898

1564 CTTCACTGGGAGACATCGCTTGGGGCCCTTGTCTGAGGACACATTTGGTAACTGCCCTT 1623  
Db CTTCACTGGGAGACATCGCTTGGGGCCCTTGTCTGAGGACACATTTGGTAACTGCCCTT 1623

1899 TACCACTGGCAAACTGGTCCCTGGGGCCAGTGCATTTGAGGACACCTCAGTATCTGCTTC 1958  
Db TACCACTGGCAAACTGGTCCCTGGGGCCAGTGCATTTGAGGACACCTCAGTATCTGCTTC 1958

1624 AATGCAACCATTTGGCTGGAATGAGAAAGCCAGCTGTGTGTAGGCAATTCAGACTCGGAGA 1683  
Db AATGCAACCATTTGGCTGGAATGAGAAAGCCAGCTGTGTGTAGGCAATTCAGACTCGGAGA 1683

1959 AACAACTACGACTTGGAAATGGGAGGCTCTCTGCTCTGTCGGCATGCAGACAGAAAA 2018  
Db AACAACTACGACTTGGAAATGGGAGGCTCTCTGCTCTGTCGGCATGCAGACAGAAAA 2018

1684 GTCTTCTGTGTCAAGAGTCACTGGGAGCAAGTAAGTACCAAAAGATGTCCAGATTTACT 1743  
Db GTCTTCTGTGTCAAGAGTCACTGGGAGCAAGTAAGTACCAAAAGATGTCCAGATTTACT 1743

2019 GTCACTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCCAAAAAATGTCTGAAAGCCCTT 2078  
Db GTCACTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCCAAAAAATGTCTGAAAGCCCTT 2078

1744 CGACCTGAAACTGTGCGCCCTGTTTCTCCCATGCAAAAAAGACTGTATTTGACTGTCT 1803  
Db CGACCTGAAACTGTGCGCCCTGTTTCTCCCATGCAAAAAAGACTGTATTTGACTGTCT 1803

2079 CGACCTGAAACTGTAGGCTTGTCTGCTTCTTGTGAAGAGCTGTATTTGACCCCA 2138  
Db CGACCTGAAACTGTAGGCTTGTCTGCTTCTTGTGAAGAGCTGTATTTGACCCCA 2138

1804 TTCAGTGAAGTGGACACCTTCCCAAGGATGTGCAAGCAGGAAATGCCACAGT---AAAA 1860  
Db TTCAGTGAAGTGGACACCTTCCCAAGGATGTGCAAGCAGGAAATGCCACAGT---AAAA 1860

2139 TATAGTGAAGTGGACATCATGCCCTCTCTGTTGAAAGAGGAGCTCCAGTATCAGGAAG 2198  
Db TATAGTGAAGTGGACATCATGCCCTCTCTGTTGAAAGAGGAGCTCCAGTATCAGGAAG 2198

1861 CAGTCTCGATACAGATCATCTCAAGAGAGCAGCAATGGAGGCCAGGAATGCCCAGAT 1920  
Db CAGTCTCGATACAGATCATCTCAAGAGAGCAGCAATGGAGGCCAGGAATGCCCAGAT 1920

2199 CAGTCTAGGATCGGGTCACTATTCAGTGTCCAGCCACCGGGGCCGAGACTGTGCACAGAT 2258  
Db CAGTCTAGGATCGGGTCACTATTCAGTGTCCAGCCACCGGGGCCGAGACTGTGCACAGAT 2258

1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTTCTGTTCTGTATATCGTGGGAAG 1980  
Db ACCTTATATGAGGAGAGAGTGTGAAGATGTTTCTTCTGTTCTGTATATCGTGGGAAG 1980

2259 CCCCTCTATGAAGAGAGGCTGTGAGGACCTCAAGCGTGTCCAAAGCTTACAGGTGGAAG 2318  
Db CCCCTCTATGAAGAGAGGCTGTGAGGACCTCAAGCGTGTCCAAAGCTTACAGGTGGAAG 2318

1981 CCACAGAAATGGAGCCCTTGATCTTGTAGTCCAGAGTGTCTGCGAGGGAATACCGGC 2040  
2319 ACTCACAATGGCGCAGATGCCAATTAGTCCCTTGGAGCGTGCAACAGACAGCCCTGGA 2378  
2041 AGCAGTGAACCTGTGGAAAGGGTTACAAACAGAGCTGTCTCATGCACTCTGTATGAC 2100  
2379 GCACAGAAAGGCTGTGGCGCTGGCGCAGCAGGCAAGAGCCATTAATTGTGCGCAAGCAT 2438  
2101 AACCGGTACAGAGAAATGATGGAATGCTCAAGCAGACAAACGGCATGCTCTCTCTTGTG 2160  
2439 GGAGGACAGCTGGAATCCATGAGTGCCTACAGTATGCGAGGCCCTGTGCGAGCCCTTACC 2498  
2161 CAGAGTGCACAGTCCCATGTGAGAGAGCTGCACCTTCACTGCTGCTGCTGCTGCTGCTG 2220  
2499 CAGGCCCTGCCAGATCCCTGCCAGGATGCTGCAATTGACACAGCTGCTGCTGCTGCTGCT 2558  
2221 CCCTGCTCCAGCACTGTGAAGCACACAAAGTAGGCGGCGACAGCTCAGAGGAAAGC 2280  
2559 TCATGCAATGAGACTGTGTGAGTGTAGGACAGAAAGCGCACTCTTGTGGAAGAT 2618  
2281 AGAAGAGAGAGAAATCCAGGATTTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
2619 AAAAAGAGGAAATGTAAATTTCCATTGTATCCCTGATGAGACTCAGTATTGT 2678  
2341 CCTTGTGATGAATTTATATCCACCTTATGGAATGCTGTCAGATTGCACTTCTCCAGAA 2400  
2679 CCCTTGTGACAAATATATGACAACTGTGGGAACTGTGCACTGTATTTTACCAGAG 2738  
2401 GGCAGAAAGGAGCCCTCACCGAGACTCGGGTCAACAGCAGACAGCAAGAAATGTGAGAA 2460  
2739 GGAAGAGTGAAGTGTGCTGGGAATGAAGTACAAGGACATCAAGGAATCGGACAA 2798  
2461 GGCCTGGCTTTTCAGCAGATAGCTTCTTCTGTATGATAAAATGGAAGACCTGTGACCCCTCC 2520  
2799 GGAATATCGTTACCAAGCAATGGCATGTGATGATCAAAATGGCAGGCTGTGGAACATCT 2858  
2521 TTCTGCAGCAGCTCTGCTTACATCAAGAAATGTGTCTATCCCTGCCATTTGATTGC 2580  
2859 AGATGTACAGCCATGTTATCATTTGAGAGGCTGTGATCATCCCTGCCCTCAGACTGC 2918  
2581 AAGTTAAGCAATGGTGTAGTGGGGTCTTGCAGTTTCATCTTTGTGGAATGAGTGA 2640  
2919 AAGCTCAGTGTGCTCAACTGCTGCGCTGCAGCAAGTCTGTGGAGTGTGTGAAG 2978  
2641 ATTCGATCCCAATGGCTTAAAGAAACCTTACATGGAGAGGACCATGTCCCAACTG 2700  
2979 GTTCGTTCTTAATGGCTGCGTGAAGAAACCATATATGGAAGAGGCTTGCCTCCCAACTG 3038  
2701 GATC-----TCAAGAAATCAGGTACATGAGCAGTCCCATGTTACAGTGAAGTCAATCAG 2754  
3039 GACCATGTCAACAGGACAGGTGTATGAGTTGTCTCCATGCCACAGTGAATGCAACCCAG 3098  
2755 TATTCCTGGTTGTAGAACACTGTCTTCATGCAAAATCAACATGAGTGAAGTGCCTG 2814  
3099 TACCTATGGGTCAACAGAGCCCTGGAGCATCTGCAAGTGAACCTTTGTGAATATGCGGGAG 3158  
2815 CGCTGTGGAGAGGAAACACAACTTACGAAATCAGATGTGT---GAATCTGCGGATGGT 2871  
3159 AACTGTGGAGAGGCGTGCAACCCGAAAGTGAATGATGATGAGAAATACAGAGATGGC 3218  
2872 GAAGGTGAGCAGTGGATGAGCAACTGTGCAACAGGATGAATTTCCCCAGAAACCCAG 2931  
3219 CCTTCTGAACATGTAGAGGATTAACCTGTGTGACCCAGAGAGATGCCCTGGGCTTAGA 3278  
2932 TCCTGTCTTATGTCTCCCAATGATGTGTCTATGCTGAGTGGGACCTTTGGAGCAAA 2991  
3279 GTGTGCAAAATACCATGCCCTGAGGATGTGTGATATCTGAATGGGGTCCATGGACCCAA 3338  
2992 TGCCCCAGATCATCGATCCCCACACAAAT---GCAGAGAAAGACTGCCACCTGCTAAGA 3048  
3339 TGTGTTTTCCTTCAATCAAGCAGTTTCCGGCAAGAGGTGAGTGTATCCATCCATCAGACAA 3398  
3049 CCATCACTGAATCAAGACTTGTGTGAAGACTCAGAGTGCAGGCTTGCCTCTCTGAAT 3108

3399 CCAGTGATGAAGGAAGATCTTGGCCCTAATGCTGTGAGAAAGAACCTGTGTAACCTGAAC 3458  
3109 GAAATATGCTTCCAGTTCAGTACAATCTAACAGATGGAGCAGATGCCAGCTGAGTGAA 3168  
3459 AAAAAGTGTCTACCACTATGATTATAATGTAAACAGACTGGAGTACATGTCACTGAGTGAG 3518  
3169 AACGACCCCTGTGGTCAAGGCTCAGGACCCGCTGTCTAAGCTGTGTGTCAGTGTGATGGC 3228  
3519 AAGCAGTTTCTGGAATGGAATTAANAACAGGATGTTGGATTCTGTTGGAAGTATGGC 3578  
3229 AAGCAGTCACTGAGCAATGTGAGCAGCATTAATTTGAGAGAGCCCGCAGAGAAATGAGC 3288  
3579 AAGTCACTGAGTGAATTAATTTGGAAGCGCTTGGCTTGGAGAAAGAACTGCGCAGATGAAC 3638  
3289 ATTCCCTCTTGGTGGAAATGCTGCTCAACTGTGCTGCTCAGGCTTCCAGGCTGTGAGCA 3348  
3639 AGTCTCTGATGTTGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3698  
3349 GAGTGTTCACAGACTGTGCGCATGGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 3408  
3699 GAAATGCTTCAACATGTGGCTCAGAGGAAATGATCCGAGAGCAACAGTGAACCCAG 3758  
3409 CCAACCCAGAGAGAGGAGCGCATGCCACACAGCTTACCCAGAGGAAACCTGCCCA 3468  
3759 CCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGAACCATGCTCAAAACCTGCCCA 3818  
3469 GTGACCCCTCTCAGCTGGGTCTTGGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3528  
3819 GTGAAGCCTTGTATCGGTGCAATATGGCCAGTGTCTCCATGCCAAGTGCAGAGGCC 3878  
3529 GACTGTGGGAGAGGATTCAGATCGCAGCCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3588  
3879 CAGTGTGAGAGGAGCAACAGAAACAGGAACTTCTTGTGTAGTGAAGTGAAGTGAAGTGAAG 3938  
3589 TCTCATGAGCTGGAAGCTGTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3638  
3939 GATGATTTACGCAAGTGTGTGATGAGAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3998  
3639 -----TCAGGACAGCATCCTGGAAGCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3693  
3999 GGTATAAAAAATATGGTTCTGAGAGAAATCCTGAGCAGCCCTTCCCGAGGTGACTGTTAT 4058  
3694 TTACAGAGATGCTCAGAGTGGAGCAGCATGTGAATTAACCTGCTGCTGCTGCTGCTGCTGCTGCT 3753  
4059 TTGAAGGACTGTCTTCTGAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4118  
3754 GAGACTGTGGCCCGCCAGTCTAGATCAAGGACTTTTATTAATTCAGTCTTTTGAAGAACAA 3813  
4119 GGCTTTGGTGAATACAGGTCAAGTCCAGACCGGTGATTAACAAGAACTAGAGAAATCAG 4178  
3814 GACAGTCCGCCCCACAGGTTCTAGAACACCGCTTGTACAGAGGCAATGTTATCAC 3873  
4179 CATCTGTGCCAGAGCAGATGTTAGAAACAAATCATGTTATGATGAGACAGTGTATGAA 4238  
3874 TACATATGGAAGAGAAAGTCTTTTGGAAACAATAACGAACGAACTGTATGCTGCGCAGCTTCA 3933  
4239 TATAAATGATGAGCGCAGTCTTTGGAAGGCTCTTCCCGAACAGTGTGTGTGTCAAAAGGTCA 4298  
3934 GATGCGTTTATGTCAAGAGGCTGTCTCCCTCAGGCCCTCTCTGCTGCTGCTGCTGCTGCTGCT 3993  
4299 GATGTTATAAATGTAAACAGGGGCTGTCTGCTGATGAGCCAGCTGATGCCAGAGTCT 4358  
3994 TGCATTCAGGCTGCAGAAAAACCTTTCTCTACTGTACACAGGGTGGAGTCTGTGGTTGT 4053  
4359 TGTAAACCAACGTTGTAGTCAACCCCACTCTGCTACTGTAGCGAGACAAACATGCCATTGT 4418  
4054 GAGAAGGCTATACAGAGATTAATGAATCAAAATGTTTCTGGAATTAATGCA----- 4105  
4419 GAAGAAGGCTACACTGAAGTCAATGCTTCTTAACAGCACCTTGAAGCAATGACACTTATC 4478  
4106 -----TGAAGTACACAGGCTCAGAGGATTAAGAGCTGATGTGAAACCTTCTCTG-G 4157



Db 4479 CCGTGGTGGTATTACCCACCATGGAGGACAAAGAGAGATGTGAACACCATCGGGCT 4538  
Qy 4158 GAAAAACAGACCTGTGAATCAAAAAATACATGATATTTTAAAGGA--TGCTCTTTCAA 4215  
Db 4539 GTACATCAACACCAACCTCCAGTAACCCAGCAGGACGGGAAAGGACCTGGTTTCTACAG 4598  
Qy 4216 CCATTGTATCCAGATGGCGGAGTAAATAATTTGGGTTTATGGCGTTTTCAGGTGGCGCTTTT 4275  
Db 4599 CCATTGGCCAGATGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGCAATTT 4658  
Qy 4276 CTCATCATGATTTTCTTAATATTTACTTCTTACCTGTGTTTGAAGAAGCCA---AAACCA 4332  
Db 4659 GTGTTACTCATCTTTATGTCTCCATGATTTTATCTAGCTTGCAAAAAAGCCAAAGAAACCC 4718  
Qy 4333 CATCAACACACACCTCCCAACAGAGGCTCTGACCTTTAGCCTACGATGAGACTTAGAC 4392  
Db 4719 CAAGAAGGCAAAACACCGACTGAACCTTTAACCTTAGCTATGATGAGATGCCGAC 4778  
Qy 4393 ATGTAA 4398  
Db 4779 ATGTAA 4784

## RESULT 2

US-10-074-566-59  
; Sequence 59, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 59  
; LENGTH: 5669  
; TYPE: DNA  
; ORGANISM: human  
US-10-074-566-59

Query Match 28.0%; Score 1232.8; DB 15; Length 5669;  
Best Local Similarity 57.7%; Pred.No. 0;  
Matches 2598; Conservative  
10 TGCAATTCAGAGCTGACCGAATCTGGTTGCAAAATGAATATGCAACACTTTCCTT 69

Db 14 TGCAATTCAGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 73  
Qy 70 CAGCCTCCTCAGACAGAGCTTGCCTCATTCCTTGTCCCGGATTTGTAGTATCTGAG 129  
Db 74 AAGCCTCTCCTGGAGCAGCTTGCCTCATTCCTTGTCCAGAGATTTGCATCGTGTCTGA 133  
Qy 130 TTCTTACCATGCTCCAACTGTAGCAAGGATGTGGGAAGAAATTCAGCATAGAACTGCG 189  
Db 134 TTTTCTGCTGTGCTCGAATGCTCCAAAGACCTCGCGCAGCGGCTCCAGCACCGGACGGT 193  
Qy 190 GGGGTATAGCTCCCGCTCTCTTTGGTGTTCGAAATGTCGAAATCTGACTGAGTCAAGA 249  
Db 194 CATGGTGGCGCCCCCGCAGTTCGGAGGCTCTGGCTGTCCAACTGACGGAGTTCAG 253  
Qy 250 GCTGTGATGCTCCCATTTCTGTCTCTTGGGGAAGAGATATACATTTAGCCTTTAAG 309  
Db 254 GTGTG-----CCAATCCAGTCCATGCGAGGCGGAGGAGCTCAGGTACAGCCTGCAT 304  
Qy 310 GTTGACCATGGAGTAATGCACTGCTCATCTTAAAGAAATTAATCCAAAGCGGAGA 369  
Db 305 GTGGGCGCTGGAGCAGCTGCTCAATGCCCACTCCGACCAAGTAAGACAAGCAAGGAGA 364  
Qy 370 ACTGTTCTGCAATTTTAA-----CTCTGATTTCAAATGAG 402  
Db 365 CGCGGAAGAAATAAGAACCGGAAAGGACCCGACGAAAGGAGTAAGAGATCCAGAGCC 424  
Qy 403 CGAGTCACCTTTTAAACATCAAAAGTTTACA-----AAGCACATCATCTTCGAAGTCT 453  
Db 425 CGCGAGCTTTATTAAGAAAAGAGAAACAGAAACAGGACAGACAGAAACAAATAT 484  
Qy 454 TGGGCAATAGATAGGTTTATCAAAACCGGAGGTTTCTGTGTACAAAGATGATGACAA 513  
Db 485 TGGGACATCCAGATTGGATATCAGACCAGAGAGTTATGTGCATTAAACAGACGGGAAA 544  
Qy 514 AATGCTATCTTAAAGCCTTTTCCCTTCAAGATTCCTTCCCATGACTGTTCAGTCTGATC 573  
Db 545 GCTGCTGATTTAAGCTTTTCCAGCAAGAGAAAGCTTCCAATGACCTTCCAGTCTGTGTG 604  
Qy 574 ATGCCAAAGACTGTGAACCTCCAGAGGTCCTCTCGAGCGCCCTGCTCCAAAGCATGC 633  
Db 605 ATCACCAGAGTGCAGGTTTCCGAGTGTGTCAGAGTGGGCCCCCTGCTCAAAAACATGC 664  
Qy 634 CGTTCAGGGAGTCTCTTGGCAGGATTTAGGAGCAGGAGCCGGAACGTGAAGCACATGGT 693  
Db 665 CATGACATGCTGCTCCCTCGAGCACTCGTGTGAAGGACACGAAACCATCAGGCAGTTCC 724  
Qy 694 ATTGGAGGTGGAAGAGAGTGTCTGAACTTCTTGAGAAAGAGGCTGCAATTTGTGAAGA 753  
Db 725 ATTGGCAGTGAAAGAGAGTGTCCAGAAATTTGAAGAAAAGAACCCCTGTTTGTCTCAGGA 784  
Qy 754 GAACTTCTGCAGCAATGTCCTGAGAACTTCTGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 785 GATGGAGTTGTCCCTGTGCCAGTATGCTGGAGAACTACAGAGTGGAGTGAAGTCCCT 844  
Qy 814 GTCTCTCTCTCTCTCCTCCTCAGAGCAGAGATCCCACTGCTGAGCGGACCCGCTGTGTC 873  
Db 845 GTGGACCCCTTTGCTCAGTCAGCAGGACAAAGAGCGCGGCAACCAAGACCGCCCTCTCTGGA 904  
Qy 874 GGTGGGATCCAGACCCGAGGAGTGTACTGTGCCAG-----AGCGTACCAGCA 921  
Db 905 GGGGGCATCCAGACCCGAGAGTGTACTGGTGCAGGCCAACGAAACCTCTCTCACAA 964  
Qy 922 GCTGCCGCACTGAGGGCCCAAGAGAGTCTCTAGACCTGTGTGAAAAGGCATTATGTGTGGGA 981  
Db 965 TTAAGTACCACCAAGAAACAAAGAGGCTCAAGCCAAATGAACTTAAATTTATCACTGGA 1024  
Qy 982 CCGCCCGCTTGGCTCTCAGCTCTGCAATATCCCTTGTCTACGGACTGCAAGTATCT 1041  
Db 1025 CCTATCCTTAATACTACAGAGCTGTGCCAATTCCTTTGCAACTGAATGTGAGTTCA 1084  
Qy 1042 TCCTGGTGCAGCTGGGCGCTGTGCATCTCCATGAAAACCTGTCAAGAACTCAGGCGGAAAAA 1101  
Db 1085 CTTTGGTGCAGCTGGGAGACCTTGTACTTATGAAAACTGTAATGATCAGCAAGGAAAAA 1144

1102 GGATTTAGACGAGGCGG-----CCATGTCTCATGGAATCTACAGGGCTGCA 1152  
1145 GGCTTCAACTGAGGAAGCGGCGCATTAACAATGAGCCCACTGGAGCTCTGGGGTAACC 1204  
1153 GGGCAATTGCCCTCATTTGTGGAGTCTGTTCCTTTGTAGGATCCAAATGTCTACCCATGG 1212  
1205 GGAATCGCCCTCACATTACTTGAAGCAATTCCTGTGAAGAGCCTCCCTGTATGACTGG 1264  
1213 ---CTGCATCAGAGGATCTGTTCCCTGTATCATGGA---AATGTGCCCTGGACAT 1266  
1265 AAAGCGGTGAGACTGGGAGACTCGAGCCAGATAACGGAAGGAGTGTGTCAGGCACG 1324  
1267 CGTATTCTGAAGCGGCTCTCCAGATGACCGCGGAGAGATGTATCAGGGAGTCTTTTC 1326  
1325 CAAGTTCAAGAGGTGTGTGATCAACAGTGAAGAGAGAGTTCAGACAGAGCTGTGC 1384  
1327 CCAGTTCCTCCCTCTCTCTGAGAGAGAGTCTTTGAAATTCCTTCGCCAATGAGACTGTGTG 1386  
1385 AGAGATGCCATCTTCCCATCCCTGTGGCTGTGATGCCCAATGCCGGAAGACTGTGTG 1444  
1387 CTGAGCGAGTGGAGGAGTGTATCTGTCATCTGTTCCAGTCCGTGTTCAATAAATAACTCAGAT 1446  
1445 CTGACGACATGGTCTAGCTGGTCTCTCTGCTCTACACCTGCTCAGGGAAGACAGAA 1504  
1447 GGGAAACAGACACAGGTCAAGAACTATCTCTGSC---ACTGGCTGGGAAGGTGGAAGCCA 1503  
1505 GGGAAACAGATACAGACAGATCCATCTGGCTATGCGGGTGAAGAGGTGGAATTCG 1564  
1504 TGTCCCTCTAGTCAAGTCTCTCAGAGCATGCTTTGTGTAATCACTATCCCTGTATGCGAG 1563  
1565 TGTCCAAATAGCAGTGTCTTTGCAAGAGTACGAAGCTGTAAATGAGCATCTTGCACAGTG 1624  
1564 CTTCACTGGGAGACATCGCTTGGGGCCCTTTGTTCTGAGACACATTTGTAATCGCCCTT 1623  
1525 TACCCTGGCAACTGTGCTCCCTGGGCCAGTGCAATTGAGGACACCTCAGTATCGTCTTC 1684  
1624 AATGCAACCAATTGGCTGGAAATGGAAGAGCAAGTGTGGTGTAGGCATTCAGATCCGAGAG 1683  
1685 AACACAACTACGATTGGAATGGGAGGCGCTCTGCTCTGCGCATGCGACAGCAAGAAA 1744  
1684 GTCTTCTGTCTCAAGAGTCAAGTGGGCAAGTAATGACCAAAAGATGTCAGATTTCTACT 1743  
1745 GTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAATGCTCTGAAAGCCCTT 1804  
1744 CGACCTGAACCTGTGCGCCCTGTTTCTCCCAATGCAAAAGAGTGTATGTGACTGTCT 1803  
1805 CGACCTGAACCTGTGAAAGGCTGTGCTGCTCTTGTAAAGAGAGTGTATTTGTGACCCCA 1864  
1804 TTCAAGTGGAGACACCTGCGCAAGGATGTCGCAAGCAGGAATGCCAGT---AAAA 1860  
1865 TATAGTACTGGACATCATGCCCCCTCTTGTGTAAAGAGGGAGTCCAGTATCAGGAAG 1924  
1861 CAGTCTCGATACAGAAATCATCATCAAGAGCAGCCAAATGAGGCCAGGAATCCAGAT 1920  
1925 CAGTCTAGGCATCGGTCATCATTCAGCTGCGACCAACGCGGGCCGAGACTGCACAGAT 1984  
1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTCTTGTGCTGTATATCGTGGAG 1980  
1985 CCCCTCTATGAAGAGAGGCGCTGTGAGGACCTCAAGCGTGCCAAAGCTACAGGTGGAAG 2044  
1981 CCAGAGAAATGGAGCCCTTGATCTTATGTGCGAGAGTCTGTCTGCGAGGGAATAAGGGGC 2040  
2045 ACTCAGAAATGGCCAGATGCCAATTAGTCCCTTGGAGCGGTGCAAGACAGCCCTTGA 2104  
2041 AGCAGTGAAGCTGTGGAAGGGGTTTCAAAACAGAGCTGTCTCATGCACTCTGTATGAC 2100  
2105 GCACAGGAGGCTGTGGGCTTGGGCGAGCAGGCAAGAGCCATTACTTCTGCAAGACAGAT 2164  
2101 AACCGGTACAGAGAAATGATGGAATGCTCTCAAGCAGACAAACGCAATGCTCTCTTTGTG 2160  
2165 GGAGCAGAGGTGGAATCCATGAGTGTCTACAGTATGACAGCCCTGTGCGAGCCCTTACC 2224

2161 CAAGAATGSCACAGTCCCATGTGCGAAGAGACTGCACCTTCACTGCTTGTGCTCAAGTTACG 2220  
2225 CAGGCTGCCAGATCCCTCTGCCAGATGACTGTCAAATTGACCAGCTGGTCCAGTTTCT 2284  
2221 CCCTGCTCCACCAACTGTGAAGCCACAAAAGTAGGCGGCGACAGCTCACAGGGAAGC 2280  
2285 TCATGCAATGGAGACTGTGTGTCAGTTAGGACCAAGAGGCGACTCTTGTGTGAAAAGT 2344  
2281 AGAAGAGGAGGAATGCCAGGATCTTCACTTACCTCTAGTGGAGACAGAACTATGT 2340  
2345 AAAAAGAGGAAAATGTAAAAATTTCCCATTTGTATCCCTGATGAGACTCAGTATGT 2404  
2341 CTTGTGATGAATTTATATCCCAACTTATGGAACCTGTCAGATTGCATTTCTTCCAGAA 2400  
2405 CTTGTGCAAAATATATGCAACACCTGTGGGAACCTGTCAGACTGTATTTTACCAGAG 2464  
2401 GGCAGAGGAGGAGCTCACCGAGGAGTGGGTACAACGACAGACAAAGATGGAAGAA 2460  
2465 GGAAGAGTGAAGTGTGCTGGGAATGAAAGTACAAGGAGACATCAAGGAATGCGACAA 2524  
2461 GGCCTGCGCTTTTCAGAGCAGTAGCCTGTTCTGATAAAAATGGAAGACCTGTTGACCCCTCC 2520  
2525 GGAATATCGTTACCAAGCAATGCGATGTACGATCAAAATGSCAGGCTTGTGGAACATCT 2584  
2521 TTCTGCAGCAGCTCTGTTTACATTCAGAAAAAATGTGTCAATCCCTGCCCATTTGATGTC 2580  
2585 AGATGTAAACAGCCATGTTTACATTCAGAGGAGCCGTCATCTCCCTGCCCTCAGACTGC 2644  
2581 AAGTTAAGCGGATGCTGTAGTTGGGGTCTTTCAGTTCATCTTGTGAAATTTGGAGTGAGA 2640  
2645 AAGCTCAGTGTGTTCCAACTGTGCGCTGCGAGCAAGTCTGTGGAGTGTGTGAAG 2704  
2641 APTGCATCCAAATGGCTAAAGAAAAAACCCTTACAAATGGAAGACACCATGTCCTCCAACTG 2700  
2705 GTTCTTCTTAAATGGCTGCGTGAATAACCATATATATGGAAGAGGCGCTTGGCCCAACTG 2764  
2701 GATC-----TCAGAATCAGGTACATGAGGAGTCCCATGTTACAGTGTGCAATCAG 2754  
2765 GACCATGTCAACAGGCGACAGGTGTATGAGTTGTCCCATGCCACAGTGTGCAACAG 2824  
2755 TATTCCTGGGTGTAGAACACTGTTCTTCATGCAAAATCAACAAATGAGCTGAGTCCCTG 2814  
2825 TACCTATGGGTACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATCGGGAG 2884  
2815 CGCTGTGAGAGGAGNACAATCTAGGAAATCAGATGTGT---GAATACCTCGCGATGCT 2871  
2885 AACTGTGAGAGGCGGTGCAACCCGAAAGTGAAGTGCATGCAAGATACAGCAGATGCG 2944  
2872 GAAGGTGAGCAGTGGATAGCAACCTGTGCAACAGGATGAAATTCCTCCAGAAACCCAG 2931  
2945 CTTTCTGAACATGTAGAGGATTAACCTCTGTGACCCAGAGAGATGCCCTGGGCTCTAGA 3004  
2932 TCCTGTTCTTATGTGTCCCAATGAGTGTGTCAATGTGAGTGGGAGCTTTGGAGCAAA 2991  
3005 GTGTGCAAAATACCATGCTGAGGAGTGTGTGATATCTGAATGGGTCTCATGACCCAA 3064  
2992 TGCCCAACAGTCAATGCGATCCCAACAAAT---GACAGAGAAGAACTCGCCACCTGCTAGA 3048  
3065 TGTGTTTGCCTTGCATCAAGCAGTTCGCGCAAGGTGAGTGTGATCCCATCAGACAA 3124  
3049 CCATCACTGAATCAAGGACTTGTCTGAAGACTCAGAGGTGCGAGCTTGCCTCTCTGAAT 3108  
3125 CCAGCTGATGAGGAAGATCTTGCCTTAATGCTGTGTGAAAAGAACCTGTAAACCTGAAAC 3184  
3109 GAAATTTGCTTCCAGTTCAGTACAAATTAACAGAGTGGACACATGCGAGTCAAGTCAAA 3168  
3185 AAAAATGCTTACCATGATTTATGTAACAGTGGAGTACATGTCAGCTGAGTGAG 3244  
3169 AACGCAACCTGTGTCAAGGCGTCAAGGCCCGCCCTGCTGAAGCTGTGTGAGTGAATGGC 3228  
3245 AAGGCAATTTGTGGAATGGAATAAAAACAAGGATGTTGGATTTGTGTCGAAGTGAATGGC 3304  
3229 AAGCCAGTCAAGCATGGACCAATGTGAGCAGCATTAATTTGGAGAGGCCCCAGAGAAATGAGC 3288

Db 3305 AAGTCAGTTGACCTGAAATATTTGTAAGCGCTTGGCTGGAGAAGAACTGCGAGATGAAC 3364  
Qy 3289 ATTCCCTGCTTGGTGAATGGTGTCAACTGTGAGTCTCAGGTTGAGCGGCTTGGACA 3348  
Db 3365 ACGTCTGATGCTGGTGAATGCCCTGTGAACGTGTGAGCTTTCTGATGCTTCCCTTGTGCA 3424  
Qy 3349 GAGTGTTCACAGACCTGTGGCCATGGAGGTGCAATGAGCCGAGTCTGATTTATCATTTATG 3408  
Db 3425 GAATGTTCTCAACATGTGGCTCACAGGAAATATGATCCGAAAGCAACAGTCAACCCAG 3484  
Qy 3409 CCAACCCAGGAGAGGAGCGCCATGCCACAGAGCTTACCCAGGAGAAACCTGCCCA 3468  
Db 3485 CCCTTTCAAGGTGATGGAAGACCATGCGCTTCCCTGATGAGACCACTGCAAAACCTTGCCCA 3544  
Qy 3469 GTGACCCCTGCTACAGCTGGTCCCTTGGCAACTGCTGCACTGTAAATTTGGAGGGTGA 3528  
Db 3545 GTGAGACCTTGTATCGGTGGCAATATGCGCAGTGTCTCCATGCCAAGTCCAGAGGCC 3604  
Qy 3529 GACTGTGGGAAGAGTTCAGATCCGCAAGCTTTCCCTGCATGGTCCACAGTGGTTCAATA 3588  
Db 3605 CAGTGTGGAAGAGGACCAAGAACAAAGGAACATTTCTTGTAGTAAGTGAATGGTCACT 3664  
Qy 3589 TCTCATGACGTGGAGCTGTGAGGATGCACTGTGTGAGAAATGCCCTT----- 3638  
Db 3665 GATGATTTCAAGAAAGTGTGATGAGAAATCTGTGTGATGAACTGAACTCATTTAGAT 3724  
Qy 3639 -----TCAGGACAGCATCCTGAAGCAGCTGTGTCTGTGCGCTTCCCGAGGAGTGCAT 3693  
Db 3725 GGTAAATAAAATATGTTCTGAGGAATCCTGACCCAGCCTTGGCCAGGTGACTGTTAT 3784  
Qy 3694 TTAACAGATGCTCAGAGTGGAGCATGTGAATTAACCTGCATGATGGAAGAGCTTT 3753  
Db 3785 TTGAAGACTGTGTTCTCCGAGCCTGTGTGAGTCACTGTGTGAATGGTGAAGATCTA 3844  
Qy 3754 GAGACTGTGGCGCCGAGTCTAGATCAAGAGTCTTTAATTAATTCAGTCTTTTGAAGACAA 3813  
Db 3845 GGTGTTGGTGAATACAGTTCAGATCCAGACCGGTGATTAATACAGAACTAGAGAACTCAG 3904  
Qy 3814 GACAGTGGCCCAACAGTCTTAGAAACACGCCCTTGTACAGGAGGCAATGTTATCAC 3873  
Db 3905 CATCTGTGCCCAAGAGCAGATGTTAGAAACAAATCATGTTATGATGGAAGTGTATGAA 3964  
Qy 3874 TACACATGGAAGCAAGTCTTTGGAAACATAACGAACGAATGATGTTGCGCAGGTTCA 3933  
Db 3965 TATAATGATGAGCGAGTCTTGGAGGCTCTTCCGAAAGTGTGTGTGTCAGAGTCA 4024  
Qy 3934 GATGGCTTAATGTACAGGAGCTGCTCCCTCAGGCGGCTGCTGCTGCCAATTCGGCAG 3993  
Db 4025 GATGTTATAAATGTAACAGGGGCTGCTTGGTGTGATGAGCCAGCTGATGCCAGAGTCT 4084  
Qy 3994 TGCAATTCAGCTGCAAGAAACCTTTCTCCTACTGTACACAGGGTGGAGTCTGTGTTGT 4053  
Db 4085 TGTAAACCCAGCTGTAGTCAACCCCACTCGTACTGTGTGAGGAGCAAAACATGCCATTGT 4144  
Qy 4054 GAGAAGGGTATACAGAGATAATGAATCAAAATGTTTCCCTGGATTTACTGCA----- 4105  
Db 4145 GAAGAAGGTGTACTAGTCAATGCTTTCTAAGCAGCACCCTTGGCAATGCACTTATC 4204  
Qy 4106 -----TGAAATGACAGCTCAGAGGATTAAGAGTGTGTAAGAAACCTTTCTG-G 4157  
Db 4205 CCGTGGTGGTATTACCCCACTGGAGGACAAAGAGGAGATGTGAAACCAAGTCCGGCT 4264  
Qy 4158 GAAAAACAGACCTGTGAATTCAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
Db 4265 GTACATCCAAACCAACCTTCCAGTACCCAGAGCGGAGGAGACCTGTTTCTACAG 4324  
Qy 4216 CCATTTGATCCAGATGCCGAGTAAATAATTTGGTTTATGGCTTTTCAAGTGGCGCTTTT 4275  
Db 4325 CCATTTGGGCGAGATGGAGACTTAAAGACCTGGTTTACGGTTGTAGCAGCTGGGCAATT 4384  
Qy 4276 CTATCATGATTTTCCATAATTTACTTCTTCTACTCTGTTTGTGAGAGGCA---AAACCA 4332

Db 4385 GTGTACTCATCTTTTATTTGTCTCCATGATTTATCTAGCTTGCATAAAGCCAAAGAACCC 4444  
Qy 4333 CATCAAGACACACCTCCCAACAGAAAGCTCTGACCTTAGCTTAGCTAGCTAGCTAGAC 4392  
Db 4445 CAAAGAGGCAAAACAAACCGACTGAAACCTTTAACCTTAGCTATGATGGAGATCCGAC 4504  
Qy 4393 ATGTAA 4398  
Db 4505 ATGTAA 4510  
  
RESULT 3  
US-10-000-512-1  
; Sequence 1, Application US/10000512  
; Publication No. US20020164699A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Fernandes, Elma  
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME  
; FILE REFERENCE: 15966-556  
; CURRENT APPLICATION NUMBER: US/10/000,512  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 6373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (235)..(4998)  
; NAME/KEY: misc\_feature  
; LOCATION: (6349)  
; OTHER INFORMATION: Wherein n is a or t or g or c.  
US-10-000-512-1  
  
Query Match 28.0%; Score 1232.8; DB 13; Length 6373;  
Best Local Similarity 57.7%; Pred. No. 0;  
Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;  
  
Qy 10 TGCATTGAGAGCTGAACCGAAGTGTGGTTGCAATATGAATATGCAACACTTTGCGCTT 69  
Db 505 TGATCCAGAAAGCAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGCCC 564  
Qy 70 CAGCTTCCTACAGAACAGGCTTGCTTCCTTGTCCCGGGATTTGTAGTATCTGAG 129  
Db 565 AAGCTCTCCTGAGCAGGCTTGCCTCATTCCTTGCAGCAAGATTGCAATGCTGTA 624  
Qy 130 TTCTTACATGGTCCAACTGTAGCAAGGGATGTGGAAAGAAATTCAGCATAGAACTCGC 189  
Db 625 TTTTCTGCTTGTCCGATGCTCCAGAGCTGCGGCGCTCCAGCACCGGACGCGT 684  
Qy 190 GCGGTCAAGTCCCTCTCTTTTGTGTTGCAATGTCCTCAATCTGACTGACTGAGTCA 249  
Db 685 CATGTGTGGCGCCCGCAGTTCGAGGCTCTGGCTGTCCAAACCTGACGAGTTCAG 744  
Qy 250 GCCTGTGATGCTCCATTTCTCTCTTGGGGAAGAGGAATATACATTTAGCCTTAAG 309  
Db 745 GTGTG-----CCATCCAGTCCATGCGAGCGGAGGAGCTCAGGTACACCTTGCAT 795  
Qy 310 GTTGGACCATGAGTAAATGCAAGTCCCTCATCTTTAAAGAAATTAATCCAAAGCGGAAGA 369  
Db 796 GTGGGGCCCTGGAGCAGCTCTCTCAATGCCCACTCCGACAGTAAAGCAAGCAAGGAGA 855  
Qy 370 ACTGTTCTGAGTTTAA-----CTCTGATTTCAATGAG 402  
Db 856 CGCGGGAAGATTAAGAAACGGGAAAGGACCGGAGCAAGAGGATTAAGGATCCAGAGCC 915  
Qy 403 CGAGTCACCTTTAAACATCAAGTTACA-----AAGCACATCATCATTCGAAAGTCT 453

Db 916 CGGAGCTTATTAGAAAAAGAGAAACAGAAACAGGACAGAACAGACAGACAGAAATAT 975  
Qy 454 TGGCAATACAGATAGGTATCAAAACCGCAGGTTCGTGTGTAACAAGAGTGTATGACAA 513  
Db 976 TGGACATCAGATTGATATCAGACCAGAGAGGTATGTGATTAACAGACGGGAAA 1035  
Qy 514 AATGCTATGTAAGCCTTTGCCCTTCAAGATTCCTTCCATGTAGTGTTCAGTCCCTCATC 573  
Db 1036 GCTGCTGATTAAAGCTTTTGCACAGAGAAAGCTTCCATGACCTTCCAGTCCCTGTG 1095  
Qy 574 ATGCCCAAGACATGTGAACCTCCAGTGTCTCTGAGAGCCCTGCTCCACAGACATGC 633  
Db 1096 ATCACCAGAGTCCAGTTTCCGAGTGTGACAGTGGAGCCCTGCTCAAAAACATGC 1155  
Qy 634 CGTTCAGGAGTCTTTGCCAGGATTTAGAGCAGGAGCCGGAACGTGAAGCACATGGCT 693  
Db 1156 CATGACATGCTGCCCTCGCAGGACCTCGTGTAGGACACGAACCATCAGGAGTTTCCC 1215  
Qy 694 ATTGGAGTGGAAAGAGTGTCTGAACTTCTTGGAGAAAGAGCCCTGATTTGAAGGA 753  
Db 1216 ATTGGAGTGAAGAGGAGTGTCCAGAAATTTGAAGAAAAGAACCTGTGTTCTCAAGGA 1275  
Qy 754 GAACCTTCGACGAATGCCAGTATTCCTGGAGAACTTCTGAATGGAAGAAATGCCAA 813  
Db 1276 GATGGAGTTGTCCTCGCCACGATATGCTTGGAGAACTACAGAGTGGACTGATGCCGT 1335  
Qy 814 GTCTCTCTCTCTGAGCAGCAGATCCCACTGGCATGTGACGGAAACCGGTGTGGC 873  
Db 1336 GTGACCCCTTTGCTCACTCAGCAGGACAGAGGCGCGCAACAGACGGCCCTCTGTGA 1395  
Qy 874 GGTGGATCCAGACCCGGAGGTGACTGTGCCAG-----AGCTACAGCA 921  
Db 1396 GGGGCACTCAGACCCGAGAGTGTCTGGTGAGGCCAAGAAACCTCTCTCACAA 1455  
Qy 922 GCTGCCCACTGAGGGCAAGAGTCTCTAGACCTGTGGAAGGCAATATGTGGGA 981  
Db 1456 TTAAGTACCACAAAGAACAAAGAGCCTCAAGCCATGACTTAAATTTATGCACGTGA 1515  
Qy 982 CCGCCCGGTGCCCTCTAGCTCTGCAATATCCCTTGCTCTAGGAGTGCATGATATCT 1041  
Db 1516 CCTATCCCTAATATACACAGCTGTGCCACATTCCTTGTCCAACTGAATGTGAAGTTCA 1575  
Qy 1042 TCCGTGTGAGCTGGGCGCTGTGCATCCATGAACCTGTCATGAACCTCAGGCGGAAAAA 1101  
Db 1576 CTTTGGTCACTTGGGACCTTGTACTTATGAAGAACTGTATGATCAGCAGGAAAAA 1635  
Qy 1102 GGATTTAGAACGAGCGAGCG-----CCATGTCTCATGGAATCTACAGGCGCTGCA 1152  
Db 1636 GGCCTCAAACTGAGAAAGCGCGCATTAACCAATGAGCCCACTGGAGCTCTGGGGTAACC 1695  
Qy 1153 GGGCATTTGCCCTCATTTGGTGGAGTCTGTTCCCTGTGAGGATCCATGTCTACCGATGG 1212  
Db 1696 GGAACCTGCCCTCACTTACTGGAAGCAATTCCTGTGAAGAGCCTGCTGTATGACTGG 1755  
Qy 1213 ---CTGGCATCAGAGGATCTGTTTCCCTGATCATGGA---AATGTGCCCTGGGACAT 1266  
Db 1756 AAAGCGGTGAGACTGGGAGACTGGAGCCAGATAACGAAAGAGTGTGTCCAGGCAAG 1815  
Qy 1267 CGTATCTGAAGGCGCTCTGCCAATGACCGCGGAGAAATGATATCAGGAGTCTTTGC 1326  
Db 1816 CAAGTTCAAGAGGTTGTGTGATCAACAGTGAAGGAGAAAGTTGACAGACAGCTGTGC 1875  
Qy 1327 CCAGTTCCCTCTCTGAGAGGAGTCTTGTGAATTCCTGCCGAATGGACTGTGTG 1386  
Db 1876 AGAGATGCATCTTCCCATCCCTGTGGCTGTGATGCCCCATGCCCGAAGACTGTGTG 1935  
Qy 1387 CTGACCGAGTGGACGAGTGTCTATCCCTGTTCAGTCCAGTCTGTTCAAATAAAAACTCAGAT 1446  
Db 1936 CTCAGCACATGCTACGTGCTCTCTGCTCACACACCTGCTCAGGGAATAACAGACAA 1995  
Qy 1447 GGGAAACAGACCAGTCAAGAACTATCTCTGC-----ACTGGCTGGGAGGTGGAAGCCA 1503

Db 1996 GGGAAACAGATACGAGCAGATCCATTCTGCCCTATCGGGTGAAGAGTGGAAATGCG 2055  
Qy 1504 TGTCCCTCTAGTCAGGCTCTCCAAAGACATCGTTTGTGTATATGACCATTCCTGTATGCAG 1563  
Db 2056 TGTCAAATACAGTGTCTTTCNAGAGTACGAAGCTGTATGAGCATCTTTCACAGTG 2115  
Qy 1564 CTTCACTGGAGACATCGCTTGGGGCCCTTGTCTGAGGACACATTGGTAACCTGCCCTT 1623  
Db 2116 TACCACCTGCAAACTGTGTCCTGGGGCCAGTGTGATTGAGGACACCTCAGTATCGTCTTC 2175  
Qy 1624 AATGCAACCAATTGGCTGGAATGGAGAACCCAGCTGTGTGTAGGACATTCAGACTCGGAGA 1683  
Db 2176 AACCAACTACGACTTGGAAATGGGAGSCCTCTCTGCTGTGCGCATGCACAGAAAA 2235  
Qy 1684 GTCTTCTGTCTCAAGAGTACGTGGGACAAATAAGACCAAAAGATGTCAGATTTCTACT 1743  
Db 2236 GTCATCTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCCAAAAAATGCTCTGAAAGCCTT 2295  
Qy 1744 CGACCTGAAACTGTGGGCCCTGTTTCTCCATGCAAAAAGACTGTATTGTGACTGCT 1803  
Db 2296 CGACCTGAAACTGTAAAGCCTTGTCTGCTTCTTGTAAAGAGACTGTATTGTGACCCCA 2355  
Qy 1804 TTAGTGTGAGTGGACACCTGCCCAGGATGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
Db 2356 TATAGTGTGACATCATGCCCTCTTCGTGTAAAGAGGGACTCCAGTATCAGGAAG 2415  
Qy 1861 CAGTCTCGATACAGATCATCATCAAGAACAGCAGCAATGAGGCGCAGGAATGCCAGAT 1920  
Db 2416 CAGTCTAGGCATCGGGTCACTCATTCAGCTGCCCAACCGGGGCCGAGACTGCACAGAT 2475  
Qy 1921 ACCTTATATAGGAGAGAGTGTGAAGATTTCTTGTGTCTGTATATCGGTGGAAG 1980  
Db 2476 CCCCTCTATGAGAGAGGCTGTGAGCACTCAAGCTGCCAAGACTCAGGTGGAAG 2535  
Qy 1981 CCACAGAAATGAGCCCTTGCACTTAGTGCCAGAGTCTGTCTGGCAGGGAATAACCGGC 2040  
Db 2536 ACTCACAATAGCGCAGATGCCAATTAGTCTCTGGAGCGTGCAACAGACAGCCCTGGA 2595  
Qy 2041 AGCAGTGAAGCCTGTGGAAGGGGTACAAACAGAGCTGTCTCATGATCTCTGATGAC 2100  
Db 2596 GCACAGGAAGGCTGTGGGCCCTGGGCGACAGCAGAGCCATTACTTTGCGCAAGCAAGAT 2655  
Qy 2101 AACCGGTGACAGAAATGATGAATGCTCAAGCAGACAAAACGGCATGCTCTCTCTGTG 2160  
Db 2656 GGAGACAGGCTGGAATCCATGAGTGCCTACAGTATGAGGCCCTGTGCCAGCCCTTACC 2715  
Qy 2161 CAAGAAATGACAGTCCCATGTGCGAAGACTGTGCACCTTCACTGTGTGTGTCAGATTACG 2220  
Db 2716 CAGGCTGCGCAGATCCCTGCGCAGGATGACTGTCAATTGACCCAGCTGTGTCCAAAGTTTCT 2775  
Qy 2221 CCCTGCTCCAGAACTGTGAAGCCACAAAAAGTAGCGGCGACAGCTCACAGGGAAGAGC 2280  
Db 2776 TCATGCAATGAGACTGTGGTGCAGTTAGGACCAAGAACGCACTCTTTGTGGAAGAAAGT 2835  
Qy 2281 AGAAAGAGGAGAAATGCCAGGATTTGACCTTTACCTTACCTCTAGTGGAGACAGAACTATGT 2340  
Db 2836 AAAAAGAGGAAAAATGAAAAATTTCCCAATTTGATCCCTGATTTGAGACTCAGTATTGT 2895  
Qy 2341 CTTCTGTGATGATTTATATCCCAACTTATGGAACCTGTGAGATGCTCAGATTGCTTCCAGAA 2400  
Db 2896 CTTTGTGAATAATATAATGCAACCTGTGGGAACTGGTCAAGACTGTATTTTACAGAG 2955  
Qy 2401 GGCAGAGGAGCCCTCACCGAGGACTGCGGGTCAAGCAGACAGCAAAAGAAATGTGGAGAA 2460  
Db 2956 GGAAGTGGAGTCTTGTGGGATGAAGTACAGGAGACATCAAGAAATCGGACAA 3015  
Qy 2461 GGCCTGGCTTTGAGAGAGTACCTGTTCTGATAAAAATGGAAGACCTGTTGACCCCTCC 2520  
Db 3016 GGATATCGTTTACCAAGCAATGGCATGTACGATCAAAATGGCAGGCTTTGTGAAACATCT 3075  
Qy 2521 TTCTCAGCAGCTGTGTTACATCAAGAAAAATGTGTCTTCCCTGCCCATTTGATGTC 2580  
Db 3076 AGATGTAAACAGCCATGGTTTACATTGAGGAGGCGCTGCATCCCTGCTCCCTCAGACTGC 3135

QY 2581 AAGTTAAGCGATTGGTCTAGTTGGGGTCTTGCAGTTCACTTTGTGGAATTGGAGTGA 2640  
DB 3136 AAGCTCAGTGAAGTCAACTGCTGGCTCGAGCAAGTCCCTGGGAGTGGTGAAG 3195  
QY 2641 ATTGATCCAAATGGCTAAAGAAAACCTTCAATGGAGGACGACATGTCCTCAACTG 2700  
DB 3196 GTTCGTTCTAATGGCTGGCTGGAAGAACCAATATAATGGAGGAAGGCTTCCGCCAACTG 3255  
QY 2701 GATC-----TCAAGATCAGGTACATGAGGAGTCCCATGTTACAGTGAGTGCATCAG 2754  
DB 3256 GACATGTCACACAGGACAGGTGTATGAGGTGTTCCCATGCCCACAGTGAATGCACCCAG 3315  
QY 2755 TATTCCTGGGTTGAGAACACTGGTCTTCAATGCAAAAATCAACAATGAGTGAAGTCCCTG 2814  
DB 3316 TACCTATGGGTACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGGAG 3375  
QY 2815 CGCTGTGAGGAGGAACACAACTAGGAATAATCAGATGTGT---GAATACTCGGATGGT 2871  
DB 3376 AACTGTGAGAGGGGTGCAAAACCCGAAAGTGAATGATGCAATGAGCAATCAGCAGATGGC 3435  
QY 2872 GAAGGTGAGCAGTGGATGAGCAACCTGTGCAACAGGATGAAATTCCTCCAGAAACCCAG 2931  
DB 3436 CTTCTGAACATGTAGAGGATTAACCTGTGACCCAGAGAGATGCCCTGGGCTCTAGA 3495  
QY 2932 TCCTGTTCTTATGTGTCCTCAATGAGTGTGTCTGATGCTGAGTGGGACTTTGGAGCAAA 2991  
DB 3496 GTGTGCAAAATACCATGCGCTGGAGACTGTGTGATATCTGAATGGGTCCATGGACCCAA 3555  
QY 2992 TGCCACAGTCAATGCAATCCCAACACAAAT---GCAGAGAAAGAACTCGCCACCTGCTAAGA 3048  
DB 3556 TGTGTTTTCCTTCAATCAAGAGATTTCCGGCAAGGTGAGTCCATCAGACAA 3615  
QY 3049 CCATCAGTGAATCAGAGCTTGTGCAAGACTCAAGGTGAGCCTTGGCTCTGCAAT 3108  
DB 3616 CCAGTGTAGAGGAAGATCTTGGCCTAATCTGTGAGAAAGAACCTGTAACTGAAAC 3675  
QY 3109 GAAATTCCTTCAGTCCAGTCAATCAACAGAGTGGACACATGCGAGCTCAGTGAA 3168  
DB 3676 AAAAAGTCTACCACTATGATTAATGATACAGACTGGAGTATCATGTCACTGAGTGAG 3735  
QY 3169 AACGCACCTGTGTCAGGGGTGAGGACCCGCTGCTAAGCTGTGTGTCAGTGATGGC 3228  
DB 3736 AAGGCAGTTTGTGGAATAGGAATAAAACAAGAGTGTGGATTTGTGTCGAAGTGATGGC 3795  
QY 3229 AAGCCAGTCAGATGACCAATGTGAGCAGCAGATAATTTGGAGAGACCCAGAGATGAGC 3288  
DB 3796 AAGTCAGTTGACCTGAAATATTGTGAAGCGCTTGGCTTGGAGAGAACTGGCAGATGAAC 3855  
QY 3289 ATTCCCTGCTTGGTGGAAATGCGTGTCAACTGTCAAGCTCTCAGGGTGGACGGCTTGGACA 3348  
DB 3856 ACGTCTGTCATGGTGGAAATGCCCTGTGAACGTGACGCTTTCTGATTTGCTCTCTTGGTCA 3915  
QY 3349 GAGTGTTCAGACACTGTGGCCATGGAGGTGGAATGACCGGACTCGATTTATCATTTATG 3408  
DB 3916 GAATGTTCTCAAAACATGTGGCTCTCAGGAAATAATGATCCGAAGACGAAACAGTACCCAG 3975  
QY 3409 CCAACCCAGGAGAGGAGCGCCATGCCCCCAGAGCTTACCCAGGAGAAACCTTGCCCA 3468  
DB 3976 CCTTTCAAGGTGATGGAAGACCATGCTCCCTGATGGACCACTCCAAACCTTGCCCA 4035  
QY 3469 GTGACCCCTGTCAGAGTGGGTCTTGGCAACTGGTCTGCAATGTAATTTGGAGGGTGA 3528  
DB 4036 GTGAAGCTTGTATTGCGTGGCAATATGCGCAGTGGTCTCCATGTCGAAGTGCAGAGGCC 4095  
QY 3529 GACTGTGGGAGGAGTTCAGATCCGAGCCTTCTCTGATGGTCCACAGTGGTTCATA 3588  
DB 4096 CAGTGTGGAGAGGAGGACCAAGAACATGTTCTTGTGTAGTGAAGTGGTGCAGCT 4155  
QY 3589 TCTATGACGCTGGAGTGTGAGGATGCACTGTGTGGAGAAATCCCTT----- 3638  
DB 4156 GATGATTTTCAGCAAGTGGTGGATGAGGAATTTCTGTGCTGACATTTGAATCATATAGAT 4215

QY 3639 -----TCAGACAGCATCCTGAAGCAGCTGTGTCTGTGCTTCCCGAGAGAGCTGCCAT 3693  
DB 4216 GGTATAAAAAATATGGTTCTGGAGAAATCCTGCAGCCAGCCCTTCCCGAGTGAATTTAT 4275  
QY 3694 TTAAAGAAATGTCAGAGTGGAGCAGATGTGAATTAACCTCATTGATGGAAGAGCTTT 3753  
DB 4276 TTGAAGAACTGTCCTTCTGGAGCCTGTGTGAGTGAATTTGATGAGTGAATTTGAGATCTA 4335  
QY 3754 GAGACTGTGGGCGCCAGTCTAGATCAAGGACTTTTATAATTTCAAGTCTTTTGAAGACCA 3813  
DB 4336 GCTTTGTTGGAATAACAGTCCAGATCCAGACCGGTGATTATACAAGAACTAGAGAAATCAG 4395  
QY 3814 GACACTGCCCCCAACAGGTTCTAGAAACAGCCCTGTACAGGAGGCAATTTATATCAC 3873  
DB 4396 CATCTGTGCCAGAGCAGATGTTAGAAACAAATCAUATTTATGATGAGCAGTGTCTATGA 4455  
QY 3874 TACATATGAAAGCAAGTCTTTTGAACCAATTAACGAAACGAACTGTATGTTGCCAGCTTCA 3933  
DB 4456 TATAATGATGCCAGTGTCTTGAAGGGCTCTTCCCGAACAGTGTGTGTCAAGAGTCA 4515  
QY 3934 GATGGGTTAATGTACAGAGGCTGTCTCCCTCAGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 3993  
DB 4516 GATGTTATAAATGTAACAGGGGCTGTCTTGTGATGAGCCAGCCTGATGCCAGAGGTCT 4575  
QY 3994 TGCATTTCAGCCTGCAGAAACCTTTCTCTACTGTACACAGGGTGGAGTCTGTGTTGT 4053  
DB 4576 TGTAAACCACTGTGTGATCAACCCACTGCTACTGTAGCGAGACAAAACATGCCATTTGT 4635  
QY 4054 GAGAAGGCTATACAGAGATTAATGAATCAAAATGTTTCTGGATTACTGA----- 4105  
DB 4636 GAAGAAGGGTACACTGAAGTCAATGCTTCTTAACAGCACCTTTGAGCAATGACACTTATC 4695  
QY 4106 -----TGAAAGTACCAGGCTCAGAGATAAAAGCTGATGTGAAAACCTTTCTG-G 4157  
DB 4696 CCGGTGTTGTTATTCACCACTGAGGACAAAAGAGAGATGTGAAAACCAAGTGGGCT 4755  
QY 4158 GAAAACAGACCTGTGAATTCAAAATACATGATATTTTAAAGGA--TGFTCTCTTCAA 4215  
DB 4756 GTATATCCAAACCCCTCCAGTAAACCCAGCAGCGGGAAGGACCTGGTTTCTACAG 4815  
QY 4216 CCATTTGATCAGATGGCGAGTAAATTTGGGTTTATGGGTTTACGGTTTACGGTGGCGCTTT 4275  
DB 4816 CCATTTGGGCGAGTGGGAGACTAAGACCTGGGTTTACGGTGTAGCAGCTGGGCAITTT 4875  
QY 4276 CTCATCATGTTTCTAATATTTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4332  
DB 4876 GTGTACTCATCTTATTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 4935  
QY 4333 CATCAAAGCACACCTCCCAACAGAAAGCTCTGACCTTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 4392  
DB 4936 CAAAGAAGGCAAAACAAACCGACTGAAACCTTTAACCTTAGCTATGATGAGATGCCGAC 4995  
QY 4393 ATGTAA 4398  
DB 4996 ATGTAA 5001

## RESULT 4

US-10-074-566-1

; Sequence 1, Application US/10074566  
; Publication NO. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytak, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1

CURRENT APPLICATION NUMBER: US/10/074,566  
CURRENT FILING DATE: 2002-02-13  
PRIOR APPLICATION NUMBER: 09/619,252  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 60/144,722  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/167,785  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: 60/276,994  
PRIOR FILING DATE: 2001-03-19  
PRIOR APPLICATION NUMBER: 60/280,898  
PRIOR FILING DATE: 2001-04-02  
PRIOR APPLICATION NUMBER: 60/332,241  
PRIOR FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: 60/288,062  
PRIOR FILING DATE: 2001-05-02  
PRIOR APPLICATION NUMBER: 60/291,766  
PRIOR FILING DATE: 2001-05-17  
PRIOR APPLICATION NUMBER: 60/314,007  
PRIOR FILING DATE: 2001-06-21  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 1  
LENGTH: 6373  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (6349)  
OTHER INFORMATION: Wherein N is A, or T, or C, or G.

Query Match		28.08;	Score 1232.8;	DB 15;	Length 6373;
Best Local Similarity		57.78;	Pred. No. 0;	Mismatches 1782;	Indels 126; Gaps 17;
Matches 2598;		Conservative	0;		
Qy	10	TGCATTGAGAGCTGAACCGAAGTGTGGTTCGAAATGAAATATGCGAACAATTGTCCTT	69		
Db	505	TGCATCCAGAAAGCAAGACATTCCTGCGAGGATATCATCTGTGAGTACTTTGAGCCC	564		
Qy	70	CAGCTCTACAGAACAGGCTTGCTTCATTCCTGTCGCCGGAATGTGTAGTATCTGAG	129		
Db	565	AAGCTCTCTCGAGAGGCTTGCTTCATTCCTGCGAGGATTCGATCGTGTCTGAA	624		
Qy	130	TTCTTACATGCTCCAACTGTAGCAAGGATGTGGGAAGAAATGCAAGATAGAACTCGC	189		
Db	625	TTTTCTGCTGCTCCGAATGCTCCAGACCTCGCGAGCGGCTCCAGCACCGAGCGGT	684		
Qy	190	CGGCTCATAGTCCCTCTCTTTTGGTGTGCAATGTCCAAATCTGACTGAGTCAAGA	249		
Db	685	CATGTGTGGCGCCCGGCAATTCGAGGCTCTGGCTGTCCAAACCTGACGGAGTTCAG	744		
Qy	250	GCCTGTGATGCTCCCAATTCCTGCTCTTGGGGAAGAGGAATATACATTTAGCCTTAG	309		
Db	745	GTGTG-----CCAAATCAGTCCATGCGAGCGGAGGCTCAGGTACAGCCTGCA	795		
Qy	310	GTGGACCATGAGTAATGAGACTGCTCATCTTAAAGAAATTAATCCAGCGGAAGA	369		
Db	796	GTGGGCGCTGAGCACTGCTCAATGCCCACTCCCGACAGTAAGACAAAGGAGA	855		
Qy	370	ACTGTTCTGATTTTAA-----CTCTGATCAATGAG	402		
Db	856	CGCGGGAAGATAAAGAACGGGAAGGACCGCAGCAAGAGGTAAGAGGATCCAGAGCC	915		
Qy	403	CGAGTCCACCTTTAAACATCAAGATTACA-----AAGCACATCATCTTGAAGTCT	453		
Db	916	CGCAGCTTATTAAGAAAGAGAAACAGAAACAGGCAAGACAGACAGACAAATAT	975		
Qy	454	TGGCAATAGAGTAGTATCAAAACCGGAGGTTTGGTGTCAAGAAAGTGTAGACAA	513		
Db	976	TGGGACATCCAGATTGATATCAGCAGAGAGGTTATGTGCATTAACAGACGGGAAA	1035		
Qy	514	AATGCTATGTTAAGCCCTTGCTTCAGATTCCTTCCCATGACTGTTCACTCTGCATC	573		

Db	1036	GCTGCTGATTTAAAGCTTTTGGCAGCAAGAGAAGCTTCCAAATGACCTTCCAGTCTGTGTG	1095		
Qy	574	ATGCCCAAGAGCTGTGAACCTTCCAGTGGTCTCTCGAGCCCTGCTCAAGACATGC	633		
Db	1096	ATCACAAAGAGTGCAGGTTTCCAGTGGTTCAGAGTGGAGCCCTGCTCAAAACATGC	1155		
Qy	634	CGTTCAGGAGTCTCTTCCAGGATTTAGGAGCAGGAGCCGGAACGTGAAGACATGCT	693		
Db	1156	CATGACATGTTGCTCCCTGCGGCACTCGTGTAAAGCAACGAACCATCAGGAGTTCC	1215		
Qy	694	ATTGAGGTGGAAGAGTGTCTGAACCTTCTGAGAAAGAGGCTGCATTTGTAAGA	753		
Db	1216	ATTGAGGTGGAAGAGTGTCCAGATTTGAAGAAAAGAACCTTGTGTCTCAAGA	1275		
Qy	754	GAACCTCTGCAAGATGTCCAGGATTTCTGGAGAACTTCTGAATGGAAGAATGCCAA	813		
Db	1276	GATGAGTGTCCCTGTGCCACGTATGCTGGAGAACTACAGAGTGCATGAGTCCGT	1335		
Qy	814	GTCCTCTCTCTCGAGCAGAGATCCCACTGGCATGTGACGGGACCCGTGTGTGC	873		
Db	1336	GTGGACCTTTGCTCAGTCAGCAGGACAAAGAGCGCGCAACCAAGACGCGCTCTGTGA	1395		
Qy	874	GGTGGATCCAGACCCGAGGTGTACTGTGCCAG-----AGCGTACCAGCA	921		
Db	1396	GGGGCATCCAGACCCGAGAGGTGTACTGCTGAGGCCCAAGAAAACCTCTCTCAAA	1455		
Qy	922	GCTGCCCACTGAGGCGCAAGAAAGTCTCTAGACCTGTGGAAGAGGCAATATGTGTGA	981		
Db	1456	TTAAGTACCCACAGAAACAAAGAGGCTCAAGGCAATGGACTTAAATATATGCACTGA	1515		
Qy	982	CCGCGCCCTGCTCTCAGCTCTGCAATATCCCTGCTCTAGGAGTGCATAGTATCT	1041		
Db	1516	CTATTCCTTAATATACACAGCTGTGCCACATTCCTTGTCCAATGAAATGGAAGTTCA	1575		
Qy	1042	TCCTGCTCAGCTGGGCGCTGTGCATCCATGAAACTGTCTATGAACCTCAGGGGAAAA	1101		
Db	1576	CCTTGTGCTGAGCTGGGACCTTGTACTTATGAAACTGTATGATCAGCAAGGAAAA	1635		
Qy	1102	GGATTAGAACGAGGACGG-----CATGCTCTCATGGATCTACAGGCGCTGCA	1152		
Db	1636	GGCTTCAAACTGAGGAAGCGGCGCATTTACCAATGAGCCCACTGGAGGCTCTGGGGTA	1695		
Qy	1153	GGGCAATGCTCTCATTTGGTGGAGTCTTCTTGTGAGGATCCAAATGCTGCTACCGATGG	1212		
Db	1696	GGAACTGCTCTCTCTTACTGGAAGCAATTCCTGTGAGAGCTGCTGTATGACTGG	1755		
Qy	1213	---CTGGCATCAGAAAGGATCTGTTTCCCTGATCATGAA---AATGCGCTGGGACAT	1266		
Db	1756	AAAGCGGTGAGACTGGGAGACTGGGAGCCAGATAACGGAAGAGTGTGTCAGGACG	1815		
Qy	1267	CGTATTTGGAAGCGCTGCGCAGAAATGACCGGAGAGATGTATCAGGAGTCTTTGC	1326		
Db	1816	CAAGTTCAAGAGTGTGTGATCAACAGTGTGGAAGAAAGTTGACAGACAGTGTGC	1875		
Qy	1327	CCAGTTCCCTCTCTCTGAGAGGAAGTCTTGTGAAATTCCTGCGCAATGGAGTGTGTG	1386		
Db	1876	AGAGATGCCATCTTCCCATCCCTGTGCGCTGTGATGCCCATGCCGAAAGACTGTGTG	1935		
Qy	1387	CTGAGCAGTGAAGGAGTGTGATCTCTGTTCCAGCTCTGTTCAATAAATAAATCAGAT	1446		
Db	1936	CTCAGCATGCTTACGTGTCTCTGCTCTCAGACCTCTCAGGGAACCGACAGAA	1995		
Qy	1447	GGGAACAGACCCAGCTCAAGAACTATCTGGC---ACTGGCTGGGAGGTGGAAGCA	1503		
Db	1996	GGGAACAGATACAGACAGATCCATCTGCGCTATGCGGTGAAGAGGTGGAATTCG	2055		
Qy	1504	TGTCCCTCTAGTCAAGGATCTCTCAAGAGCATCGTTTGTGTATGACCATTCCTGTATG	1563		
Db	2056	TGTCATAATAGCAGTGTCTTTCAGAAAGTACGAAGCTGTATGAGCATCTCTTGACAGT	2115		
Qy	1564	CTTCACTGGGAGACATCGCTTGGGCGCTTCTCTGAGGACACATTTGTAACCTGCTT	1623		

Db 2116 TACCACTGCAAACTGGTCCCTCGGGCCAGTGCATGTAGGACACCTCAGTATCGTCCTTC 2175  
Qy 1624 AATGCAACCAATTTGGCTGGAATGGAGAGCCAGTGTGTGTAGGCATTCAGACTCGAGA 1683  
Db 2176 AACACAACTACGACTTGGATGGGAGGCCCTCTCTCTGTCGGCATGCACAAAGAAA 2235  
Qy 1684 GTCTTCTGTGTCAAGAGTCACTGTGGGACAGTAATGACCAAAAGATGTCCAGTCTTACT 1743  
Db 2236 GTCACTGTGTGTGCGAGTCAATGTGGGCCCAAGTGGGACCCCAAAAATGTCTCTGAAAGCCTT 2295  
Qy 1744 CGACCTGAAACTGTGCGCCCTGTCTTCTCCATGCCAAAAGAGCTGTATTGTGACTGCT 1803  
Db 2296 CGACCTGAAACTGTAAAGCCTGTCTGCTCTCTGTAAAGAGGACTGTATTGTGACCCCA 2355  
Qy 1804 TTCAGTGTAGTGAACCTCTGCCCAAGAGTGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
Db 2356 TATAGTGTAGTGAACATCATGCTCCCTCTTCTGTAAAGAGGAGGACTCCAGTATCAGGAAG 2415  
Qy 1861 CAGTCTCGATACAGAAATCATCATCAAGAGCAGACCAATGAGGCCAGGAATGCCAGAT 1920  
Db 2416 CAGTCTAGGCATCGGGTCAATCAATCAGTGCAGCCCAACGGGGCCGAGACTGCAACAGAT 2475  
Qy 1921 ACCTTATATGAGGAGAGAGTGTGAAGATGTTCTTGTCTCTGTATATCGGTGGAAG 1980  
Db 2476 CCCCTCTATGAAGAGAGAGCCCTGTGAGGCACCTCAAGCGTGCMAAGCTACAGGTGGAAG 2535  
Qy 1981 CCACGAATGAGCCCTTGATCTTAGTGCCAGAGTCTGTCTGCGCAGGAAATACGGGC 2040  
Db 2536 ACTCAAAATGCGCAGATGCGCAATAGTCTCCCTGGAGCGTGCMAAGCAGCGCCCTGGA 2595  
Qy 2041 AGCAGTGAAGCCTGTGGAAGAGGGGTACAAAACAGAGCTGTCTCATGCACTCTCTGATGAC 2100  
Db 2596 GCACAGGAGGCTGTGGCCCTGGCGCAGCAGGCAAGCCATTACTTTGTCGAACGAGAT 2655  
Qy 2101 AACCGTCAAGCAAAATGATGAAATGCTCTCAAGAGACAAAGCGCATGCCCTCTCTGTTG 2160  
Db 2656 GGAGGACAGGCTGGAATCCATGAGTGCCTTACAGTATGAGGCCCTGTGCCAGCCCTTACC 2715  
Qy 2161 CAAGAATGACAGTCCCATGTCCGAGAAGACTGCACCTTCACTGTGTGTGTCGAAGTTTACG 2220  
Db 2716 CAGGCTCGCAGATCCCTGCGCAGAGTACTGTCAATGACAGCTGTGTCCAGTTTCT 2775  
Qy 2221 CCTGTGTCAGAACTGTGAAGCCACAAAGATGAGCGCGCAGACTCACAGGGAAGAC 2280  
Db 2776 TCATGCAATGAGAGACTGTGGTCACTTAGCACCAAGAGCGCACTCTTGTGGAAGAAAGT 2835  
Qy 2281 AGAAGAGAGGAAATGCCAGGATTCAGACTTTACCTTTAGTGTAGGAGCAGACTATGT 2340  
Db 2836 AAAAGAGAGAAATGTAAAAATCCCATTTGTATCCCTGATGAGACTCAGTATTGT 2895  
Qy 2341 CCTTGTGATGAATTTATATCCCAACCTTATGGAAGTGTGTGAAATGCAATTCCTCCAGAA 2400  
Db 2896 CTTTGTGACAAATATAATGACAACTGTGGGGAAGTGTGACAGCTGTATTATTACAGAG 2955  
Qy 2401 GGCAGAGGAGGACCTCACGAGACTGCGGGTACAGCAGACAGCAAGAAATGTGGAA 2460  
Db 2956 GGAAGAGTGGAAAGTGTGTGGGAATGAAGTACAGGAGACATCAGGAAATGCGGCA 3015  
Qy 2461 GGCCTGCGCTTTCGAGCAGTAGCCTGTCTGATATAAATGGAAGACCTGTGACCCCTCC 2520  
Db 3016 GGATATCGTTACCAAGCAATGGCATGTACGATCAAAATGSCAGGCTGTGTGGAACATCT 3075  
Qy 2521 TTCTGAGCAGCTGTGTTACATTCAGAAATGTGTCAATTCCTGCGCCATTTGATTGC 2580  
Db 3076 AGATGTAAACAGCCATGTTTACATTTAGGAGGCGCTGCATCATCCCTGCGCCCTCAGACTGC 3135  
Qy 2581 AAGTTAAGGATTTGTCTAGTTGGGGTCTTGCAGTTCATCTTGTGGAATGGAGTGA 2640  
Db 3136 AAGCTCAGTGTAGTGTCCAACTGTGTGCGCTGCACAAAGTCTCTGTGGAGTGTGTGAAG 3195  
Qy 2641 ATTTCGATCAAAATGGCTAAAGAAACCTTACAAATGGAGGACGACCATGTCCCAAACTG 2700  
Db 3196 GTTGTGTTCTAAATGGCTGGTGAAGAAACCATATATATGAGGAGGAGCCCTTGGCCCAAACTG 3255

Qy 2701 GATC-----TCAAGAAATCAGGTATCATGAGGAGTCCCATGTTTACAGTGTGCAATCAG 2754  
Db 3256 GACCATGTCAACAGGACACAGGTGTATGAGTTGTCCCATGCCACAGTACTGCAACCAAG 3315  
Qy 2755 TATTCCTGGGTTGTAGAACACTGGTCTTCATCGAAAAATCAACAATGAGCTGAGGTCCTTG 2814  
Db 3316 TACTTATGGGTCAAGAGCCCTGGAGCATCTCAAGAGTGACCTTTGTGAATATGCGGGAG 3375  
Qy 2815 CGTGTGGAGGAGAACACAATCTAGAAAAATCAGATGTGT---GAATACTCGGATGTGT 2871  
Db 3376 AACTGTGGAGAGGGCGTGCAAACCCGAAAGTGAGATGCATCGAATACACAGATGAC 3435  
Qy 2872 GAAGGTGGAGAGTGTGATAGCAACCTGTGCAACCGAGATGAAATTCCTCCCAAGAACCCAG 2931  
Db 3436 CTTTCTGAAACATGTAGAGGATTACCTCTGTGACCCCAAGAGATGCCCTGGGCTCTAGA 3495  
Qy 2932 TCTGTCTCTTATGTGTGCCAATGAGTGTCTGTCTGAGTGGGAGCTTTTGGAGCAAA 2991  
Db 3496 GTGTGCAAAATTACCATGCCCTGAGGACTGTGTGATATCTGAATGGGTCCATGGACCCAA 3555  
Qy 2992 TGCACAGTCTATGCGATCCCAACAAT---GCAGAGAAAGAACTCGCCACTGTCTAAGA 3048  
Db 3556 TGTGTTTTCCTTGCATCAAAAGCAGTTTCCGCAAGGTGAGTGCATGCCATCAGACAA 3615  
Qy 3049 CCATCACTGAACTCAAGGACTTGTCTGAGACTCACAGGTGCAGCTTGCCTCTGTGAT 3108  
Db 3616 CCAGCTGATGAAGAAAGATCTTGGCCCTAATGCTTGTGAAAGAAACCTGTAACTTGAAC 3675  
Qy 3109 GAAAAATGCTTCCAGTTCAGATCAATCTAAACAGAGTGGAGCACATGCCAGCTGAGTGA 3168  
Db 3676 AAAAACTGTCTACCATATGATTAATGTAACAGACTGGAGTACATGTCACTGAGTGAG 3735  
Qy 3169 AACGACCCCTGTGTCAAGCGTCAGGACCCGCTCTAAGCTGTGTGTGCTGAGTGTGCG 3228  
Db 3736 AAGCGAGTTTGTGAAATGGAATAAAACAAGATGTTGGAATGTTTCCAAAGTGTGCG 3795  
Qy 3229 AAGCAGTCAAGATGACCAATGTGAGCAGCAATAATTTGGAGAAGCCCCAGAGAAATGAGC 3288  
Db 3796 AAGTCACTGACCTGAAATATTTGGAAGCGCTTGGCTTGGAGAGAACTGGCAGATGAAC 3855  
Qy 3289 ATTCCCTGTGTGTGAATGCGTGTGTCATCTGTGAGCTCTCAGGGTGGAGCGGCTTGGACA 3348  
Db 3856 ACCTCTGCTGATGGTGAATGCCCTGTGAACCTGTCACTTCTGATTTGGTCTCTCTTGGTCA 3915  
Qy 3349 GAGTGTTCACAGACCTGTGCGCATGAGGTGCAATGAGCGGCACTCGAATTTATCATATG 3408  
Db 3916 GAATGTTCTCAACATGTGGCTTCAAGGAAATGATCCGAAAGACAGACAGTGCACCCAG 3975  
Qy 3409 CCAACCCAAAGGAGAGGACCGGCATGCCCAACAGAGCTTACCCAGAGAAACCTTGCCCA 3468  
Db 3976 CCCTTTCAAGGTGATGGAAGACCATGCCCTTCCCTGATGGACAGTCCCAACCTTGCCCA 4035  
Qy 3469 GTACCCCTCTGTACAGCTGGGTCTTGGCAACTGGTCTGCAATGTAAATTTGAGGGTGA 3528  
Db 4036 GTGAAGCCTTGTATCGTGGCAATATGGCCAGTGTCTCCATGCCAAGTGCAGGAGGCC 4095  
Qy 3529 GACTGTGGGGAAGGAGTTCAGATTCGACAGCTTTCCTGCAATGCTCCACAGTGTGTTCAATA 3588  
Db 4096 CAGTGTGGAGAGGGACCAAGAACATTTCTTGTGTAGTAAGTGTATGGGTGAGCT 4155  
Qy 3589 TCTCATGAGCTGGAGCTGTGAGGATGCACTGTGTGGAGAAATGCCCC--- 3638  
Db 4156 GATGATTTCAAGAAAGTGTGTGATGAGAAATTTCTGTGCTGCAATTTGAATTAGAT 4215  
Qy 3639 -----TCAGACAGCATCTCGAGCAGCTGTCTCTGTGCTTGGCCAGGAGACTGCCAT 3693  
Db 4216 GGTAAATAAAATATGTTCTGGAGGATCCTCGAGCCAGCCCTTGGCCAGGTGACTGTAT 4275  
Qy 3694 TTAACAGAAATGTCAGAGTGGAGCACATGTGAATTAACCTGCAATGATGGAAGAAAGCTTT 3753  
Db 4276 TTGAAGCACTGCTCTCTCGAGGAGCTGTGTGAGTGCACCTGTGTGAATGAGTGTGAGGATCTA 4335



3754 GAGACTGTGGCGCCAGCTCTAGATCAAGGACTTTTATAAATCACTTTTGGAGAACCA 3813  
3755 GAGACTGTGGCGCCAGCTCTAGATCAAGGACTTTTATAAATCACTTTTGGAGAACCA 3814  
4336 GCGTTTGGTGGATACAGCTCAGATCCAGACCGGTGATTTATACAGAACTAGAGATCAG 4395  
3814 GACAGCTGCCCCCAACAGGTTCTAGAAACACCGCCCTTGTACAGGAGGCAATGTTATCAC 3873  
4396 CATCTGTGCCCCAGAGAGATGTTAGAAACAAAATCATGTTATGATGGACAGTGCTATGAA 4455  
3874 TACACATGAAGAGCAAGCTTTTGGACAAATAACGAACAACTGTATGGTCCAGGGTTCA 3933  
4456 TATAAATGATGGCCAGTCTCTGGAAAGGCTCTTCCGACAGTGTGGTGTCAAGAGTTCA 4515  
3934 GATGGGTTAAATGTACAGAGAGCTGCTCCCTCAGGCGCCCTGCTGCTGCCATTCGGCAG 3993  
4516 GATGGTATAAATGTAAACAGGGGCTGCTTGGTGTGATGAGCAGCCCTGATGCCAGAGTCT 4575  
3994 TGCAATCCAGCTGTCAGAAACCTTTCTCTACTGTACACAGGTTGGAGTCTGTGGTGT 4053  
4576 TGTAACCCACCGTGTAGTCAACCCACTGTTACTGTAGCGAGCAAAAACATGCCATTT 4635  
4054 GAGAAGGCTATACAGAGATAATGAATCAAAATGTTTCTCTGGATTACTGCA----- 4105  
4636 GAAGAAGGCTACACTGAAGTCATGTTCTTAACAGCACCCCTTGAGCAATGCACACTTATC 4695  
4106 -----TGAAGTACAGGCTCAGAGGATAAAGAGTGTGTAAGAACCTTTCTG-G 4157  
4696 CCGTGGTGTATTAACCCAGCTGAGGAGGAGAGATGTGAAGAACCCAGTCCGGCT 4755  
4158 GAAAGACAGACCTGTGAATTCAAAATAATACATGATATTTTAAAGGA--TGGTCTCTCAA 4215  
4756 GTACATCCAAACCAACCTCCAGTACCCAGCAGGAGCGGGAGGACCTGGTTCTACAG 4815  
4216 CCACTTGATCCAGATGGCGAGTAATAATTTGGGTTTATGGGTTTACGTTTACGTTGCGCTTT 4275  
4816 CCAATTTGGGCGAGATGGAGACTTAAAGACCTGGGTTTACGTTTACGTTTACGTTGCGCTTT 4875  
4276 CTCACTCATGATTTTCTATATTTACTTCTACTGTTTGAAGAGCA--AAACCA 4332  
4876 GTGTTACTCATCTTTATGTTCTCCATGTTTATCTAGCTTGCAAAAGCAAGAACCC 4935  
4333 CATCAAGCACACCTCCCAAGAGAGCTCTGAGCTTACGCTTACGATGAGAGCTTAGAC 4392  
4936 CAAAGAAGGCAAAACCAACCGACTGAAACCTTTAAACCTTAGCCTATGATGAGATGCCGAC 4995  
4393 ATGTAA 4398  
4996 ATGTAA 5001

## RESULT 5

JS-10-074-566-58  
Sequence 58, Application US/10074566  
Publication No. US20030207348A1  
GENERAL INFORMATION:  
APPLICANT: Shimkots, Richard A.  
APPLICANT: Fernandes, Elma R.  
APPLICANT: Li, Li  
APPLICANT: Gorman, Linda  
APPLICANT: Gusev, Vladimir Y.  
APPLICANT: Padigaru, Muraidhara  
APPLICANT: Patturajan, Meera  
APPLICANT: Shenoy, Suresh G.  
APPLICANT: Sytek, Kimberly A.  
TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
FILE REFERENCE: 15966-556 CIP1  
CURRENT APPLICATION NUMBER: US/10/074,566  
CURRENT FILING DATE: 2002-02-13  
PRIOR APPLICATION NUMBER: 09/619,252  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 60/144,722  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/167,785  
PRIOR FILING DATE: 1999-11-29

PRIOR APPLICATION NUMBER: 60/276,994  
PRIOR FILING DATE: 2001-03-19  
PRIOR APPLICATION NUMBER: 60/280,898  
PRIOR FILING DATE: 2001-04-02  
PRIOR APPLICATION NUMBER: 60/332,241  
PRIOR FILING DATE: 2001-11-14  
PRIOR APPLICATION NUMBER: 60/288,062  
PRIOR FILING DATE: 2001-05-02  
PRIOR APPLICATION NUMBER: 60/291,766  
PRIOR FILING DATE: 2001-05-17  
PRIOR APPLICATION NUMBER: 60/314,007  
PRIOR FILING DATE: 2001-08-21  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 58  
LENGTH: 6373  
TYPE: DNA  
ORGANISM: human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (6349)  
OTHER INFORMATION: Wherein N is A, or T, or C, or G.  
US-10-074-566-58  
Query Match  
Best Local Similarity 28.08; Score 1232.8; DB 15; Length 6373;  
Matches 2598; Conservative 0; Mismatches 1782; Indels 126; Gaps 17;  
QY 10 TGCATTCAGAGCTGACCGAACTGTGGTTGCAAAATATGCAATATGCGAAACACTTTGGCCCTT 69  
DB 505 TGCATTCAGAAAGCAAGACATTCCTGCGGAGGATATCATCTGTGAGTACTTTGAGCCC 564  
QY 70 CAGCCTCTACAGAACAGGCTTGCCTCATTTCTTGTCCCGGAGTGTGTAGTATCTGAG 129  
DB 565 AAGCCTCTCTGAGCAGGCTTGCCTCATTTCTTGCAGCAAGATTGTCATCTGTCTGAA 624  
QY 130 TTCTTACATGTCCTCAACTGTAGCAAGGATGTGGAGAAATTTGAGCATAGAACTCGC 189  
DB 625 TTTCTGCTGTGTCGGAATGCTCCAGACCTTGGCAGCGGCTCCAGCACCGGACGCGT 684  
QY 190 GCGGTATAGTCCCTCTCTTTTGGTGTTCGAATGTCCAAATCTGACTGAGTCAAGA 249  
DB 685 CATGTGTGGCGCCCGCAGTTCGAGGCTGTGGTGTCTCAAACTGAGGAGTTCAG 744  
QY 250 GCCTGTGATGTCCTCATTTCTCTCTTGGGAGAGGAATATACATTTAGCCTTAAG 309  
DB 745 GTGTG-----CCAACTCAGTCCATGCGGCGGAGGCTCAGGTACAGCCTGCAT 795  
QY 310 GTTGGACCATGGAGTAAATGCAGACTGCTCATCTTTAAAGAAATTAATCCAAGCGGAAGA 369  
DB 796 GTGGGCGCTGGAGCACCTGCTCAATGCCCACTCCCGACAGTAAAGACAAGCAAGGAGA 855  
QY 370 ACTGTTCTGATTTTAA-----CTCTGATTCGAATGAG 402  
DB 856 CGCGGGAAGAAATTAAGAACCGGAAAGACCGCAGCAAGAGTAAAGGATCCAGAGCC 915  
QY 403 CGAGTCACTTTAAACATCAAGTTTCA-----AAGCACATCATCATTCGAAGTCT 453  
DB 916 CGCGAGCTTATTAGAAAGAGAGAAACAGAAACAGGACAGACAGCAAGAGACAATAT 975  
QY 454 TGGCAATAGAGATAGTTTATCAAAACCGGCGAGTTCGTGTGTACAGAAAGTATGACAA 513  
DB 976 TGGGACATCCAGATTGGATATCAGACCAGAGAGTTATGTGCATTAAACAGACGGGAAA 1035  
QY 514 AATCCTATGTTAAGCCTTTCCTTCAAGATTCTTCCATTTGACTGTTTCACTGCTGCATC 573  
DB 1036 GCTGCTGATTTAAGCTTTTTCAGCAAGAGAGCTTCCAAATGACCTTCCAGTCTCTGTG 1095  
QY 574 ATGCCCAAGACTGTGAACCTCCAGTGTCTCTCTGAGAGCCCTCTCTCAAGACATGC 633  
DB 1096 ATCACCAGAGTGCAGGTTTCGAGTGTGTGAGTGGAGGCGCTCTCTCAAAACATGC 1155  
QY 634 CGTTCAGGAGTCTCTTGGCCAGGATTTAGGAGCAGGAGCGGAAAGTGAAGCATGGCT 693

1156 CATGACATGCTGCTCCCTGAGGACCTCGTGTAGGACACGACCAATCAGGAGTTTCCC 1215  
694 ATTGGAGTGGAAAGAGTGTCTGTAACCTTCTGAGAAAGAGGCTGCAATGTTGAGGA 753  
1216 ATTGGCAGTGAAGAGGAGTGTGAGAAATTTGAAGAAAGAAACCTGTTGTCTCAAGA 1275  
754 GAACTTCTGAGCAATGCTCCAGAGTATCTCTGAGAACTTCTGAATGGAAAGAAATGCCAA 813  
1276 GATGGAGTGTGCTCCCTGCTGCAAGATGCTGAGAACTTACAGAGTGGACTGAGTSCCGT 1335  
814 GTCTCTCTCTCTGAGAGCAGAGAGTCCCACTGGCATGTGACGGGAACCGTGTGTGGC 873  
1336 GTGACACCTTGTCTAGTACAGAGGAGCAAGAGGCGGGCAACAGACGGCCCTCTGTGA 1395  
874 GGTGGGATCCAGACCCGGGAGGTGTACTGTGCCAG-----AGCGTACCAGCA 921  
1396 GGGGCAATCCAGACCCGAGAGGTGTACTGTGCGGAGGCAACGAAACCTCTCTCACAA 1455  
922 GCTGCCGCACTGAGGGCAAGAGTCTCTAGACCTGTGGAAGAGGCAATATGTGGGA 981  
1456 TTAAGTACCACAGAAACAAAGAGGCTCAAGGCAATGAGCTTAAATTTATGCACTGA 1515  
982 CCCGCCGCTGCTCAGCTCTGCAATATCCCTTCTCTACGAGCTGCAATGATCT 1041  
1516 CCTATCCCTAATCTACAGAGCTGTGCCAATTTCTTGTCCAACTGAATGTGAAGTTTCA 1575  
1042 TCCTGGTCAAGCTGGGCTGTGATCCATCCATGAAGTGTATGAAGCTCAGGGGAAAAA 1101  
1576 CTTTGGTCAAGCTGGGCTGTGATTTAATGAAGCTGAATGATCAGCAAGGGAAGAAA 1635  
1102 GGATTTAGAGAGGAGCG-----CCATGCTCTCATGAACTACAGGGCCTGCA 1152  
1636 GGCTTCAAACTGAGGAGGCGGCAATTAACAAATGAGCCCACTGAGGCTCTGGGTAACC 1695  
1153 GGGCAATGCTCAATTTGGTGGAGTCTGTCTCTGTGAGGATCCAAATGCTCTACCGATGG 1212  
1696 GGAAGCTGCTCACTTACTTGGAGGCAATTTCCCTGTGAAGAGCTGCTGTATGACTGG 1755  
1213 ---CTGGATCAGAGGATGTTTCCCTGTATCATGGAA---AATGTGCTGGGACAT 1266  
1756 AAAGCGGTGAGACTGGGAGACTGGAGCCAGATAACGGAAGAGGTGTGGTCAGGCAAG 1815  
1267 CGTATTTGAAGGCGCTGTGCAAGATGACCGGAGAGAGTGTATCAGGAGTCTTTGGC 1326  
1816 CAAGTTCAAGAGTGTGTGATCAACAGTGTGAGAGAGAGTTGACACAGCTGTGC 1875  
1327 CCAGTTCCTCTCTGAGAGAGAGTCTGTGAAATTCCTGCGCAATGGAAGTGTGTG 1386  
1876 AGAGATGCCATTTCCCATCTCTGTGGCCTGTGATGCCCATGCCCCGAAAGACTGTGTG 1935  
1387 CTGAGCGAGTGGAGGAGTGTCTATCTGTTCAGTCTCCAGTCTCTCAATAAATACTCAGAT 1446  
1936 CTGAGCAGATGTTCTAGTGTCTCTCTGCTCAGACACTGCTCAGGGAAGAAACGACAGAA 1995  
1447 GGGAAACAGACCCAGGTCAAGAACTATCTGTGC---ACTGTGCTGGGAGAGTGGAAAGCCA 1503  
1996 GGGAAACAGATACGAGCAGCATCAATTTCTGCTGTGCGGTGAAAGAGGTGGAATTCGC 2055  
1504 TGTCTCTCTCTGAGGCTCTCCAGAGCATGCTTTTGTGTAATGACATTTCTGTATGTCAG 1563  
2056 TGTCCAAATAGAGTGTCTTCAAGAGAGTACGAAGCTGTGAAGTGTGATGAGTCTTGCACAGT 2115  
1564 CTTCACTGGGAGACATGCCCTTGGGGCCCTTGTCTGTGAGGACATTTGGTAACTGGCCCTT 1623  
2116 TACCACCTGGCAACTGTGTCTCTGGGCGAGTGCATTTGAGGACACTCAGTATCGTCTTTC 2175  
1624 AATGCAACCATGCTGGAATGGAGAGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1683  
2176 AACACAACTACGACTTGAATGGGAGGCTCTCTGCTGTGTGCGCATGCGACAGAGAAA 2235  
1684 GTCTTCTGTCTCAAGAGTCAAGTGGGAGCAAGTAATGACAAAGATGTCCAGATTTCTACT 1743

2236 GTCATCTGTGTGCGAGTCAATGTGGGCCAAAGTGGGACCCCAAAAATGTCTCTGAAGCCTT 2295  
1744 CGACTGAAACTGTGCGCCCTGTTTCTCCATGCAAAAAGACTGTATTTGTGACTGCT 1803  
2296 CGACTGAACTGTAAAGCCTTGTCTCTTCTTGTAAAGAGCTGTATTTGTGACCCCA 2355  
1804 TTCAGTGTGTGACACCCCTGCCCAAGGATGTGCAAGCAGGAATAATGCCAGT---AAAA 1860  
2356 TATAGTGTGAGTGTGATGCTGCTCTCTGTTGTAAGAAAGGGGACTCCAGTATCAGGAAG 2415  
1861 CAGTCTCGATACAGAAATCATCCAGAGAGCAGCAATGAGGCGCAGGAATGCCAGAT 1920  
2416 CAGTCTAGGCAATGGGTCTCATCTTCTGCTGCCCAACGCGGGCCGAGACTGTCACAGAT 2475  
1921 ACCTTATATGAGGAGAGAGAGTGTGAAGATGTTTCTTGTGTCTGTATATTCGGTGAAG 1980  
2476 CCCCTCTATGAAGAGAGGCGCTGTGAGGCACCTCAAGCGTGCACAAAGCTACAGGTGGAAG 2535  
1981 CCACAGAAATGGAGCCCTTGCATCTTGTAGTCCAGAGTCTCTCTGGCAGGGAATAACGGG 2040  
2536 ACTCACAATGGCGAGATGCCAATTAGTCCCTTGGAGCGTGCAACAGACAGCCCTTGA 2595  
2041 AGCAGTGAAGCCTGTGGAAGAGGGGTTACAAACAAAGAGCTGTCTCATGCATCTCTGTATG 2100  
2596 GCACAGAGAGCTGTGGCCTGGCGACAGCAAGAGCCATTACTTGTGCAAGCAAGAT 2655  
2101 AACCGTGTGAGAGAAATGATGGAATGCTCAAGCAGACAAACGCGCATGCTCTCTTGTG 2160  
2656 GGAGCAGAGGCTGGAATCATGAGTGTCTACAGTATGCAAGCGCTGTGCCAGCGCTTACC 2715  
2161 CAAGAATCCACAGTCCCATGTGAGAGAGCTGACCTTCACTGCTGTGCTCAAGTGTACG 2220  
2716 CAGGCTGCGAGATCCCTGCGCAGAGTGTCAATTTGACAGCTGGTCCAAGTCTTCT 2775  
2221 CCCTGCTCCACGAATGTGAAGCCACAAAAGTGTAGGCGGAGACAGCTCAGGGAAGAAC 2280  
2776 TCATGCAATGAGAGCTGTGTGTGAGTGTAGGACACAGAAAGCGCACTCTTGTGTGAAAAGT 2835  
2281 AGAAGAGAGGAGAAATGCGAGGATCTGACCTTACCTCTAGTGGAGACAGAACTATGT 2340  
2836 AAAAAGAGGAAATGTAAATTTCCATTTGTATCCCTGATGAGACTGATGTATTTACAGAG 2895  
2341 CTTGTGATGAATTTATATCCAACTTTATGGAACTGGTCAAGTTCATTTCTTCCAGAA 2400  
2896 CTTTGTGACAAATATATGTACAACTGTGGGAACTGGTCAAGTGTATTTTACAGAG 2955  
2401 GGCAGAGGAGCCTCAGGAGAGTGGGTAACAGCAGACAGCAAGAAATGTGGAGAA 2460  
2956 GGAAGAGTGGAGAGTGTGCTGGGAATGAAGTGAAGAGACATCAAGGAATGGGACAA 3015  
2461 GGCCTGCGCTTTGAGAGAGTGTGCTTGTATATAAATGGAAGACCTGTGTGACCCCTCC 2520  
3016 GGATATCGTTACCAAGCAATGCGATGTACGATCAAAATGSCAGGCTTGTGGAACAATCT 3075  
2521 TTCTGCGAGAGTGTGTTTACATTCAGAAATAATGTGTATCTCTGCGCCATTTGATGTC 2580  
3076 AGATGAACAGCCATGTTTACATTTGAGAGGCTGTGATCATCCCTGCGCTCAGACTGC 3135  
2581 AAGTTAAGCGATGTTGTGTGTTGGGGTCTTGTGAGTTCATCTGTGGAATGTGAGTGA 2640  
3136 AAGCTCAGTGTGTGCTCACTGTGCGGCTGCGAGAGTCTGTGTGGAGTGTGTGAAG 3195  
2641 ATTGATCCAAATGGCTAAAGAAACCTTACAAATGAGGAGCAGCAATGTCCCAAACTG 2700  
3196 GTTCGTCTTAATGCTGCTGGAAGAACCAATATAATGAGGAGGCGCTTGGCCCAAACTG 3255  
2701 GATC-----TCAAGAACTCAGTACATGAGCAGTCCCATGTGTACAGTGTGCAATCAG 2754  
3256 GACCATGTCAACAGGAGCAGGTGTATGAGGTGTCTCCATGCCACAGTGTGCAACAG 3315  
2755 TATTCCTGGGTTGTAGAACACTGGTCTTTTCATGCAAAATCAACAATGAGCTGAGTCCCTG 2814  
3316 TACCTATGSGTCAAGAGCCCTTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAG 3375

2815	CGCTGTGGAGGAGGAACA	CAATCTAGGAAAAATCAGATGTGT	----GAATATCTGCGATGGT	2871
3376	AACTGTGGAGAGGCGTGC	AAACCCGAAAAAGTGAGATGC	ATGCAGAAATACAGCAGATGGC	3435
2872	GAAGGTGGAGCAGTGGAT	AGCAACCTGTGTCAACACGAGAT	AAAAATCCCCCAGAAAAACCCAG	2931
3436	CCTTCTCTGAACATGT	AGAGGATTACCTCTGTGAC	CCAGAGAGATGCCCTGGGCTCTAGA	3495
2932	TCTGTGTCTTTATGTGT	CCCAATGAGTGTGTATGTCT	GAGTGGGACATTTGGAGCAAA	2991
3496	GTGTGCAAAATTACAT	TGCCCTGAGGACTGTGTATAT	CTCAANTGGGGTCCATGGAGCCCAA	3555
2992	TGCCCAGGTCTATGCCAT	CCCCACACAAT---GCAGAGAA	AAACTGCCCACTGCTTAAGA	3048
3556	TGTGTTTGGCTTTGCAAT	CAAGACGTTTCCGGCAAGGT	CAGTGCATCCCATCAGACAA	3615
3049	CCATCACTGNACTCAGG	ACTTGTGTGAAGACTCACAG	TGTGAGCTTGCCCTCTTGAAT	3108
3616	CCAGCTGATGAAGGAT	ATCTCCCTTAATGCTGTT	GAGAAAGACCCCTGTAACCTGAAC	3675
3109	GAAAAATTGCTTCCAGT	TCCAGTACAATCTTAACAG	AGTGGAGCATGCCCACTGAGTGA	3168
3676	AAAAACTGCTACCAT	TATGATTATATGTAA	CAGACTGGAGTACATGTCACTGAGTGAG	3735
3169	AAGCACAACCTGTGTGT	CAAGCGTTCAGACCCGCT	TGCTTAAGCTGTGTGTGCACTGATGGC	3228
3736	AAGCGAGTTGTGGAA	TGGAAATAAAACAAGGAT	GTGTGATTGTGTTCCAAAGTGATGGC	3795
3229	AAGCCAGTCAGCATG	AGCAATGTGAGCAGCATAA	TTTGGAGAAGCCCCCAGAGAAATGAGC	3288
3796	AAGTCAGTTGACCT	GTAAATTTGTGAAGCGCT	TGGCTTGGAGAAAGAACTGGCAGATGAAC	3855
3289	ATTCCCTGCTTGGT	GAATGCGTGGTCAACTGT	CAGCTCTCAGGGTGGACGGGTCTGGACA	3348
3856	AGCTCTGTCATGGT	GAATGCCCTGTGTGA	ACTGTCTGATTGGTCTCCTCGTCA	3915
3349	GAGTGTTCACAGAC	CTGTGGCCATGGAGTTC	GATGAGCCGACTCGATTATCATATTG	3408
3916	GAATGTTCTCAACA	TGTGGCTTCACAGGAA	AATGATCCGAAGACCAACAGTGACCCAG	3975
3409	CCAAACCAAGGAGAG	AGGACGGCCATCCCCAC	AGAGCTTACCCAGGAGAAAACTGCCCA	3468
3976	CCCTTTCAAGGTGAT	GGAGACCATGCCCTT	CCCTGATGGACCAACCCCTGCCCA	4035
3469	GTGACCCCTGCTAC	AGCTGGGTCTTGCCAACT	GGCTTGCATGTAAATTGGAGGGTGA	3528
4036	GTGAGCCTTGTTAT	TCGGTGGCAATATGGC	CACTGTCTCCATGGCAAGTGCAGGAGGC	4095
3529	GACTGTGGGGAAGG	AGTTCCAGATCCGACG	CTTCCCTGTCATGTGTCCACAGTGGTCAATA	3588
4096	CAGTGTGGAGAGGG	ACAGAACATTTCTTG	TGTAGTAGTGATGGTCAGCT	4155
3589	TCTCATGCACTGG	ACGTGCGAGGTGACTGT	GTGGAGAAATGCCCCCTT-----	3638
4156	GATGATTTCAGCA	AAAGTGGTAGAGGAAT	CTGTGCTGACATTGAATATAGAT	4215
3639	-----TCAGGA	CAGCATCCTGAAGAC	AGCTGTGTCTGTGCGCCACGAGACTGCCAT	3693
4216	GGTAATAAAAA	TATGGTTCTGGAGGAAT	CCTGCAGCCAGCTTGCCAGGTGACTTTAT	4275
3694	TTAAACAAGATG	TCAGATGGAGCACA	TATGTAATTAACCTGCAATTGATGGAGAAGCTTTT	3753
4276	TTGAAGGACTGTG	TCCTCGAGCCGTGT	CTCAGCTGACCTGTGTGTAATGGTGAGGATCTTA	4335
3754	GAGACTGTGGGCG	CCAGTCTAGATCAAG	CAATTTTATAATTCAGTCTTTTGAGAACCCAA	3813
4336	GGCTTTGGT	GGAATACAGGTCAGAT	CCAGCCGGTGTATATACAGAACTAGAGAACTCAG	4395
3814	GACAGCTGCC	CCCCAACAGGTTCT	TAGAAAACAGCCCTTTGTACAGGAGGCAAAATGTTATCAC	3873
4396	CATCTGTGCC	CAGAGCAGATGTT	TAGAAAACAAATCATGTTATGATGGACAGTCTCATGAA	4455

Qy	3974	TACACATGGAAAGCAAGTCTTTTGGAAACAATAACGAACGAACACTGTATGTGTGCCACAGCGTTCA	3933
Db	4456	TATAAAATGGATGGCCCACTGCTTCTTGGAAAGGGCTCTTCCGGAACAGTGTGGTGTCAAAGGTCA	4515
Qy	3934	GATGGCGCTTAATGTTCACAGGAGGCTGTCCCTCTCAGCCCGCTCTCTGTGCCATTCCTGGCAG	3993
Db	4516	GAATGGTATAAATGTACAGGGGGCTGCTTGGTGATGAGCCAGCTGATGCGGACAGGTCT	4575
Qy	3994	TGCATTCCAGCCTGCGAGAAACCTTTCTCCTACTGTACACAGGGTGGAGTCTCTGTGGTTGT	4053
Db	4576	TGTAACCCACCGTGTATGTCAACCCCACTCGTACTGTAGCGAGACAAACCATGCCATTGT	4635
Qy	4054	GAGAAGGGCTATACAGAGATAATGAAATCAAAATGGTTTCTCTGGATTACTGCA-----	4105
Db	4636	GAAGAAGGTACACTGAAGTCATGTCTTTACAGCACCCCTTGAGCAATGCACATTATC	4695
Qy	4106	-----TGAAGTACACAGCTCAGAGGATAAAAAAGCTGATGTGAAAACTTTCTG-G	4157
Db	4696	CCCGTGGTGGTATTATCCACCATGGAGGACAAAGAGGAGATGTGAAACCAAGCTCGGGCT	4755
Qy	4158	GAATAACAGACCTGTGAATTCAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA	4215
Db	4756	GTACATCCAAACCCCAACCTTCAGTAAACCCAGCAGGACGGGGAAGGACCTGGTTTCTACAG	4815
Qy	4216	CCACTTGTATCCAGATGGCCGAGTAAAAAATTTGGGTTTATGGCGCTTTCAGGTGGCGCTTTT	4275
Db	4816	CCATTTGGGCCAGATGGAGACTAAAGACCTGGTTTACGGTGTACAGCTGGGGCATTT	4875
Qy	4276	CTCATCATGATTTTCTTAATATTACTTCTCCTACTTGTTCGAGAGGCA---AAACCA	4332
Db	4876	GTGTTACTCATGTTTATTGTCTCCATGATTTACTAGCTTGCATAAAGGCAAAAGAACCC	4935
Qy	4333	CATCAAGCACACCTCCCAACAGAGAGCTCTGACCTTAGCCTTACCATGGAGACTTTAGAC	4392
Db	4936	CAAGAAGGCAAAACACCGACTGAACCTTTAACCTTTAGCCTATGATGGAGATGCGAC	4995
Qy	4393	ATGTAA	4398
Db	4996	ATGTAA	5001

RESULT 6  
US-10-037-270-332  
; Sequence 332, Application US/10037270  
; Publication No. US20030104529A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyun  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunging  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: Tillinghast, John  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and  
; FILE REFERENCE: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/10/037,270  
; CURRENT FILING DATE: 2002-01-04  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104

SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 332  
LENGTH: 6378  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (271)..(5037)  
NAME/KEY: misc feature  
LOCATION: (1)..(6378)  
OTHER INFORMATION: n = a,t,c or g  
JS-10-037-270-332

Query Match 28.0%; Score 1229.6; DB 14; Length 6378;  
Best Local Similarity 57.6%; Pred. No. 0;  
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;  
2Y 10 TGCAATTCAGAGCTGACCGAATGTGGTTGCAATGAATATATGCAACACTTGGCCCTT 69  
DB 541 TGCATCCAGAAAGACAAAGACATTCCTGGGAGGATATCATCTGTGAGTACTTTGAGCCC 600  
2Y 70 CAGCCTCCTACAGAACAGGCTTGCCTCATTCCTGTGCCCCGGGATTTGTAGTATCTGAG 129  
DB 601 AAGCCTCTCTGAGCAGGCTTGCCTCATTCCTGTGCCAGCAAGATTGCATCGTGTGAA 660  
2Y 130 TTCTTACATGGTCCAACTAGCAAGAGGATGCGGAAGAAATTGCAGCATGAAGTCCG 189  
DB 661 TTTTCTGCTGTGCTCGAATCTCTCAAGACCTGCGGAGCGGGCTCCAGCACCGGAGCGGT 720  
2Y 190 GCGGTCTATAGTCCCTCTCTTTTGGTGGTTTGCATGTCCAAATCTGACTGAGTCAAGA 249  
DB 721 CATGTGTGGGCCCCCGAGTTTCGAGGCTCTGGGCTGTCCAACTGTACGAGTTCCAG 780  
2Y 250 GCGTGTGATCTCCCATTTCTCTCTTGGGGAAGAGAAATATACATTTAGCCCTTAAG 309  
DB 781 GTGTG-----CCAAATCCAGTCCATGCGAGCGAGGAGCTCAGGTACAGCTGTGAT 831  
2Y 310 GTTGGACCATGGAGTAATACAGACTGCTCTCATCTTAAAGAAATTAATCCAGCGGAAGA 369  
DB 832 GTGGGGCCCTGGAGCAGCTCTCTCAATGCCCACTCCCGACAGTAAGAACAGCAAGGAGA 891  
2Y 370 ACTGTCTTGGATTTTAA-----CTCTGATTCAAATGAG 402  
DB 892 CGCGGGAAGATAAAGAACGGGAAAGAGCCGAGCAAGAGGATAAAGGATCCAGAGCC 951  
2Y 403 CGAGTCACTTTAAACATCAAGTTACA-----AAGCATCATCATTCGAGTCT 453  
DB 952 CGCGAGCTTTAAGAAAGAAAGAGAAACAGCAAGAGAGGTTATGTGCATTAAAGAGCGGAA 1011  
2Y 454 TGGCAATAGAGATAGGTATCAAAACCGGCGAGTTTCGTGTACAAGAAAGTATGACAA 513  
DB 1012 TGGACATCAGATTGGATATCAGACAGAGAGGTTATGTGCATTAAAGAGCGGAA 1071  
2Y 514 AATGCTATGTAAGCCTTTGCTTCAAGATTCCTTCCCATTTGATTTGATCTGCTGCTGATC 573  
DB 1072 GCTGCTGATTTAAGCTTTTCCAGCAAGAGAAAGCTTCCATGATCTTCCAGTCTCTGTG 1131  
2Y 574 ATGCCCAAGACTGTGAACCTCCAGTGTCTCTCTGGAGCCCTGCTCCAGACATGC 633  
DB 1132 ATACCAAGAGTCCAGGTTTCGAGTGTAGAGTGGAGCCCTGCTCAAAAACATGC 1191  
2Y 634 CGTTCAGGAGTCTCTTTGCCAGGATTTAGAGCAGGAGCGGAACTGAGACATCGGCT 693  
DB 1192 CATGACATGGTGTCCCTGCGGACCTCGTGTAAAGACACGAACCATCAGGAGTTTCCC 1251  
2Y 694 ATTGGAGTGGAAAGGATCTCTGACTTCTTGAAGAAAGAGCCGCTGATTGTGAAGA 753  
DB 1252 ATTGGCAGTGAAGAGGATGTCAGATTTTGAAGAAAGAAACCTTGTGTCTCAAGGA 1311  
2Y 754 GAATCTTCGACCAATGCTCCAGGATTTCTCTGGAGAACTTCTGAATGGAAAGAAATGCCAA 813  
DB 1312 GATGGAGTTGTCCCTGTGCCAGTATGCTGTGGAGAACTACAGAGTGGACTGAGTGGCGT 1371

QY 814 GTCTCTCTCTCTCTCTCGAGCAGCAGGATCCCACTGCGCATGTGAGCGGACCCGCTGTGTGGC 873  
DB 1372 GTGGACCTTTTGTCTAGTCAGCAGGACAGAGCGCGGCAACAGACGCGCTCTGTGGA 1431  
QY 874 GGTGGATCCAGACCCCGGAGGTGTACTGTGCCCAG-----AGCGTACAGCA 921  
DB 1432 GGGGCGATCCAGACCCCGAGAGGTGTACTGTGCGTCAGGCCAACGAAACCTCTCTCACAA 1491  
QY 922 GCTGCCGCACTGAGGGCCAAAGGAAGTCTCTAGACCTGTGAAAAGGCATTATGTGTGGGA 981  
DB 1492 TTAAGTACCACAAAGAACAAAGAGCCTCAAGCGCAATGCACTTAAATTTATGCATGGA 1551  
QY 982 CCGCCCGCTGTGCGCTCTAGCTCTGCAATATCCCTGTCTTACGAGCTCATAGTATCT 1041  
DB 1552 CCTATCCCTAATATACACAGCTGTGCAATCTCTTGTCCAACTGAATGTGAAGTTTCA 1611  
QY 1042 TCCGTGTGCTGCGGCTGTGCAATCCATGAATACTGTCATGAACCTCAGCGGGAAGAAA 1101  
DB 1612 CCTTGTGCTGAGCTTGGGACCTTGTACTTTAAACCTGTATATATGAGGAGGAAAGAAA 1671  
QY 1102 GGATTTAGAACGAGCGAGCG-----CCATGTCTCTCATGGAATCTACAGGCGCTGCA 1152  
DB 1672 GGCTTCAAACTGAGGAAGCGCGCATTTACCAATGAGCCCACTGGAGCTCTGGGTAACC 1731  
QY 1153 GGGCATTTGCCCTCATTTTGGTGGAGTCTGTTCTGTGAGGATCCAATGTCTACCGATGG 1212  
DB 1732 GGAACCTGCCCTCACTTACTGGAAGCCATTCCTCTGTGAAGGCTTGCCTGTATGACTGG 1791  
QY 1213 CTGCA---CAGAAAGGATCTGTTCCTCTGATCATGGA---AATGTGCCCTGGGACAT 1266  
DB 1792 AAAGCAGTGAGACTGGGAGACTGCGAGCCAGATAAAGGAAGAGTGTGTCTCCAGGACG 1851  
QY 1267 CGTATTTGAAGCGCTGTGCGAGATGACGGGGAAGATGATATCAGGAGTCTTTGC 1326  
DB 1852 CAAGTTCAAGAGGTTGTGTGATCAACAGAGTGTGGAAGAAAGTTGACACAGCTGTGC 1911  
QY 1327 CCAGTTCCTCTCTGAGAGAGTCTTGTGAAATTCCTGCGCAATGGAGTGTGTG 1386  
DB 1912 AGAGTGCATCTTCCCATCTCTGTGCTGTGATGCCCTGCGGAGAGTGTGTG 1971  
QY 1387 CTGAGCGAGTGGACGGAGTGTGATCTGTTCAGTCTCTGTTCAAAATAAAATCAAGAT 1446  
DB 1972 CTCAGCATGCTGTAGTGTCTCTCTGTCTCACACACCTGCTCAGGGGAAAGACAGAA 2031  
QY 1447 GGGAAACAGACCAAGTCAAGAACTATCTCTGSC---ACTGCTGGGAGGTGGAAGCCCA 1503  
DB 2032 GGGAAACAGATACAGACAGATCTCTCTGCGCTATGCGGCTGAAGAGTGTGAAATTCG 2091  
QY 1504 TGTCCCTCTAGTCAAGGCTCTCAAGAGCATCTGTTGTATGACCATTCCTCTGTATGAG 1563  
DB 2092 TGTCCAAATAGCAGTGTCTTGCMAAGATGAGAGCTGTATGAGCATCTCTGACAGTG 2151  
QY 1564 CTTCACTGGGAGCATCGCTTGGGCGCTTGTCTGAGGACATTTGGTAATGCGCCTT 1623  
DB 2152 TACCACCTGGCAACTGTGCTCTGGGCGCAGTGCATTGAGGACACCTCAGTATCGTCTTC 2211  
QY 1624 AATGCAACCAATGGCTGGAATGGAAGACCAAGTGTGCTGTAGGCAATTCAGACTCGGAGA 1683  
DB 2212 AACACACTACGATTTGGAATGGGAGCGCTCTCTGCTCTGCGCATGCGACAGAGAAA 2271  
QY 1684 GTTCTGTGTCAAGAGTCAAGTGGGCAAGATGATGACCAAGAGATGTCCAGATTTACT 1743  
DB 2272 GTCATCTGTGTGCGAGTCAATGTGGGCGCAAGTGGGACCCCAAAATTTCTCTGAAGCCTT 2331  
QY 1744 CGACTGAAACTGTGCGCCCTGTTTCTCCATGCAAAAAGACTGTATTTGACTGTCT 1803  
DB 2332 CGACTGAAACTGTGAGGCTTGTCTGCTCTTGTGAAGAGACTGTATTTGACCCCA 2391  
QY 1804 TTCAGTGTGGAACACCTGCGCAAGGATGTGCAAGCAAGAAATGCCAGT---AAAA 1860  
DB 2392 TATAGTACTGGACATCATGCCCTCTTCTGTGTAAGAGAGGACTCCAGTATCAGGAAG 2451  
QY 1861 CAGTCTCGATACAGATCATCATCCAGAGACGCAATGGAGGCCAGGAATGCCAGAT 1920



QY 4054 GAGAAGGGCTATACAGAGATAAATGAATCAATGGTTTCTCGATTACTGCA----- 4105  
Db |||||  
QY 4672 GAAGAAGGGTACACTGAAGTCATGCTTCTTAAACAGCACCCCTTGAGCAATGCACACTTATC 4731  
Db |||||  
QY 4106 -----TGAAAGTACCAGGCTCAGAGATGAAAAGCTGATGTGAAAACCTTTCTG-G 4157  
Db |||||  
QY 4732 CCCGTGTGGTATTACCACCATGAGGAGCAAAAGAGGAGATGTGAAAACCAAGTCGGGCT 4791  
Db |||||  
QY 4158 GAAAAACAGACCTGTGAATTCAMAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
Db |||||  
QY 4792 GTACATCCAACCCACCCCTCCAGTAACCCAGCAGGAGCGGGAAGGACCTGGTTTCTACAG 4851  
Db |||||  
QY 4216 CCACCTGATCCAGATGCGCGAGTAAAAATTTGGTTTATGGCGTTTCAGGTGGCGCTTTT 4275  
Db |||||  
QY 4852 CCATTTGGCGCAGATGGAGACTAAAGACCTGGGTTTACGGTGTAGCAGCTGGGGCATTT 4911  
Db |||||  
QY 4276 CTCATCATGATTTTCTTAATATTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4332  
Db |||||  
QY 4333 CATCAAGCACACCTCCCAACAGAGACCTCTGACCTTAGCTACGATGAGACTTAGAC 4392  
Db |||||  
QY 4972 CAAGAAGGCAAAACACCCAGCTGAAACCTTTAACTTAGCTATGATGAGATGCGGAC 5031  
Db |||||  
QY 4393 ATGTAA 4398  
Db |||||  
QY 5032 ATGTAA 5037  
Db |||||

## RESULT 7

US-10-117-722-332  
; Sequence 332, Application US/10117722  
; Publication No. US20030219744A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219744A1e1 Nucleic Acids and  
; FILE REFERENCE: 784C1P2BCIP  
; CURRENT APPLICATION NUMBER: US/10/117,722  
; PRIOR FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: 09/620,312  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 332  
; LENGTH: 6378  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (271)..(5037)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(6378)  
; OTHER INFORMATION: n = a,t,c or g  
US-10-117-722-332

Query Match 28.0%; Score 1229.6; DB 15; Length 6378;  
Best Local Similarity 57.6%; Pred No. 0;  
Matches 2596; Conservative 0; Mismatches 1784; Indels 126; Gaps 17;

QY 10 TGCATTCAGAGCTGAACCGCAACTGTGTTGCAATGAAATATGCAACACTTTGCCCTT 69  
Db |||||  
QY 541 TGCATCCGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTTGAGCCC 600  
Db |||||

QY 70 CAGCCTCTCAGAACAGGCTTGCTCATCTCTTGTCCCGGGATTGTGTAGTATCTGAG 129  
Db |||||  
QY 601 AAGCCTCTCTCGAGCAGGCTTGCTCATCTCTTGCAGCAAGATTGCATCGTGTCTGAA 660  
Db |||||  
QY 130 TTCCTTACCATGGTCCCACTGTAGCAAGGATGTGGGAGAAATTTGCAGCATAGAACTCGC 189  
Db |||||  
QY 661 TTTTCTGCTGGTCCGAATGTCTCAAGACCTTCCGCGCAGCGGGCTCCAGCACCGGACGGT 720  
Db |||||  
QY 190 GGGGTCTAGCTCCCTCTCTCTTTGGTGGTTTGAATGTCCAAATCTGACTGAGTCAAGA 249  
Db |||||  
QY 721 CATGTGTGGCGCCCGCAGTTCGGAGGCTCTGCTGTCCAAACCTTGACGAGTTCAG 780  
Db |||||  
QY 250 GCCTGTGATGCTCCCATTTCTGTCTCTTGGGAAGAGGATATATACATTTAGCCTTAAG 309  
Db |||||  
QY 781 GTGTG-----CCAAATCCAGTCCATGCGAGCGCGAGGAGCTCAGTACAGCTTGAT 831  
Db |||||  
QY 310 GTTGACCATGGAGTAAATGCAGACTGCTCATCTTAAAGAAATTAATCCAAAGCGGAGA 369  
Db |||||  
QY 832 GTGGGCGCTTGAGCACCTGTCTCATGCCCCACTCCCGCAAGTAAGCAAGCAAGAGA 891  
Db |||||  
QY 370 ACTGTCTGTGATTTTAA-----CTCTGATTTCAATGAG 402  
Db |||||  
QY 892 CGCGGGAAGAAATAAGAAACGGGAAAAGACCGCAGCAAGGAGTAAGGATCCGAGGCC 951  
Db |||||  
QY 403 CGAGTCACCTTTTAAACATCAAAAGTTTACA-----AAGCACATCATCTTCAAGTCT 453  
Db |||||  
QY 952 CCGGAGCTTTTATAGAAAAGAGAAAACAGAAACAGGAGAAACAGACAGAGAAACAAATAT 1011  
Db |||||  
QY 454 TGGCAATAGAGATAGTTTATCAAAACCGCGAGGTTTGTGTACAAGAAAGTGTATGGACAA 513  
Db |||||  
QY 1012 TGGGACATCCAGATTGGATATCAGACCAGAGAGTTATGTGCTATTAAACAGACGGGAAA 1071  
Db |||||  
QY 514 AATGCTATGTTAAGCTTTGCTTCAAGATTCCTCCATTTGACTGTTTCACTGCTGATC 573  
Db |||||  
QY 1072 GCTGCTGATTTAAGCTTTTGGCAGCAAGAGCTTCCATGACCTTCCAGTCTCTGTGTG 1131  
Db |||||  
QY 574 ATGCCCCAAGACTGTGAACCTCCAGTGTCTCTTGGAGCCCTGTCTCAAGACATGTC 633  
Db |||||  
QY 1132 ATCAACAAAGAGTGCAGGTTTCCAGTGTGTAGAGTGGAGCCCTGTCTCAAAACATGC 1191  
Db |||||  
QY 634 CTTTCAAGGAGTCTTCTCCAGGATTTAGGAGCAGGACCGGAGCGTGAACACATGGCT 693  
Db |||||  
QY 1192 CATGACATGTTGCCCTTGCAGGCACTCGTGTAGGACACCAACCATCAGGAGTTTCCC 1251  
Db |||||  
QY 694 ATTGAGGTGGAAGAGTGTCTGAACTTCTTGAAAGAGGCTGTGATTTGTAAGGA 753  
Db |||||  
QY 1252 ATTGAGGTGGAAGAGTGTCTGAAATTTGAAGAAAAGAACCTGTGTTCTCTCAAGGA 1311  
Db |||||  
QY 754 GAACTTCTGAGCAATGTCCAGGATTTCTTGGAGAACTTCTGAAATGAAAGAAATGCCAA 813  
Db |||||  
QY 1312 GATGAGTGTCTCCCTGTGCCACGTATGGCTGGAGAACTACAGAGTGGACTGAGTGGCGT 1371  
Db |||||  
QY 814 GTCTCTCTCTCTCGAGCAGGAGTATCCCACTGGCATGTGAGGAGCCCTGTGTGGC 873  
Db |||||  
QY 1372 GTGGACCTTTGCTCAGTCAGCAGGACAAAGAGGCGCGCAACCAAGACGCGCTCTGTGA 1431  
Db |||||  
QY 874 GGTGGATCCAGACCCGGAGGTGTCTGTGCCAG-----AGCGTACCAGCA 921  
Db |||||  
QY 1432 GGGGGCATCCAGACCCGAGAGTGTCTGCGTGCAGGCGCAACGAAACCTCTCTCAAA 1491  
Db |||||  
QY 922 GCTGCGGCTAGAGGCGCAAGAGTCTCTAGACTGTGGAAGAGGCAATATGTGTGGA 981  
Db |||||  
QY 1492 TTAAGTACCCACAAGAACAAAGAGCCCTCAAGCCAAATGGACTTAAATTTATGCACTGGA 1551  
Db |||||  
QY 982 CCGGCGCCCTTGGCTCTCAGCTCTGCAATATCCCTTGTCTTACGAGCTGCAATATCT 1041  
Db |||||  
QY 1552 CCTATCTCTTAATACACAGCTGTGCCACATTTCTTGTCCAACTGAATGTGAAGTTTCA 1611  
Db |||||  
QY 1042 TCTGTGTAGCCCTGGGCGCTGTGCATCCATGAAACTGTGTAACCTCAGACCTCAGGGGAAAAA 1101  
Db |||||  
QY 1612 CTTTGTGTAGCTTGGGACCTTGTACTTATGAAAACTGTAAATGATCAGCAAGGGAAAAA 1671  
Db |||||

1102 GGATTTAGACGAGCAGCG-----CCATGTCCTCATGAATCTACAGGCGCTGCA 1152  
1672 GGCTTCAAAGTGAAGAGCGCGCATTAACAATGAGCCACCTGGAGGCTCTGGGTAAACC 1731  
1153 GGCATTTGCCCTCATTTGGTGGAGTCTGTTCCCTGTTGAGGATCCAAATGTGCTACCGATGG 1212  
1732 GGAAACTGGCCCTCACTTACTGGAAGCAATCCCTGTGAAGAGCTGCTCTGTTATGACTGG 1791  
1213 CTGGCA--TCGAAAGGATCTGTTTCCCTGATCATGGA---AATGTGCCCTGGAGAT 1266  
1792 AAGCAGTAGAGCTGGAGAGCTCGAGCCAGATACGGAAGAGAGTGGTTCAGGCGAG 1851  
1267 CPTATTCTGAAGCCCTCTGCCAGATGACCGGGAGAGATGTATCAGGAGTCTTTGCG 1326  
1852 CRAAGTTCAAGAGGTTGTGTCATCAACAGTGTGGAAGAGAGTTGACAGACAGCTGTGC 1911  
1327 CAGTTTCCCTCTCTCTGAGAGGAGTCTTGTGAAATTCCTCGCGAATGACACTGTGTG 1386  
1912 AGAGATGCCATCTTCCCAATCCCTGTGGCTGTGATGCCCTCATGCCGAAGACTGTGTG 1971  
1387 CTGAGCGAGTGGACGAGTGGTCACTCTGTTCCAGTCTCTGTTCAAAATAAAACTCAGAT 1446  
1972 CTGAGCACATGTTCTAGTGTCTCTCTGCTCACACACTGCTCAGGGAACACGACAGNA 2031  
1447 GGGAAACAGACACAGGTCAAGAACTATCTTGGC---ACTGGCTGGGGAAGGTGGAAGCCA 1503  
2032 GGGAAACAGATACGAGCAGATCCATTTCTGGCTATGCGGGTGAAGAGGTGGAATTCGC 2091  
1504 TGTCCCTCTAGTCAGGCTCTCCAAAGAGACGTTTGTGTAATGACCATTCCTGTATGCGAG 1563  
2092 TGTCCAAATAGCAGTCTTGTGAAAGAGTACGAGCTGTAAATGAGCAGTCTTGCACAGTG 2151  
1564 CTTCACTGGGAGACATCGCTTGGGCGCTTGTCTGAGGACACATTTGTAATCTGCCCTT 1623  
2152 TACCACTGGCAACTGGTCCCTGGGCGAGTGCATTTGAGGACACCTCAGTATCGTCTTC 2211  
1624 AATGCAACCATTTGGCTGGAATGAGAGGACACGTTGTTGTTAGGCAATTCAGACTCGGAGA 1683  
2212 AACCAACTACGACTTGGAAATGGGAGGCTCTCTGCTGTGCGCATGAGACAAAGAAA 2271  
1684 GTCTTCTGTCAAGAGTCACTGGGACAAAGTAATGACCAAAAGATGTCAGATTTCTACT 1743  
2272 GTCATCTGTGCGAGTCAATGTGGCCAAAGTGGGACCCCAAAAAATGTCTGAAAGCCTT 2331  
1744 CGACCTGAAACTGTGGCGCTGTTTCTCCCATGCAAAAAAGACTGTATTTGACTGCT 1803  
2332 CGACCTGAAACTGTGAGGCTTGTCTGCTCTCTTGTGTAAGAGGACTGTATTTGACCCCA 2391  
1804 TTCAGTGAGTGGACACCCCTGCCCAAGGATGTGCCAAGCAGGAAATGCCACAGT---AAAA 1860  
2392 TATAGTCACTGGACATCATGCCCTCTTCTGTGTAAAGAGGGGACTCCAGTATCAGGAG 2451  
1861 CAGCTCTCGATACAGAAATCATCTCAAGAGAGGAGCCAAATGGAGGCGCAGAAATGCCAGAT 1920  
2452 CAGTCTAGGCATCGGCTCATCACTTCACTGCGCAGCCCAACGCGGCGCGAGACTGCACAGAT 2511  
1921 ACCTTATATGAGAGAGAGAGTGTGAAGATGTTTCTTGTGTCCTGTATATCGTGGGAG 1980  
2512 CCCTCTATGAAGAGAGGCTGTGAGGCACCTCAAGCGTGTGCCAAGCTACAGGTGGAG 2571  
1981 CCACAGAAATGGAGCCCTTGTCATCTTGTAGTCCAGAGTCTGTCTGCGAGGGAATAACGGGC 2040  
2572 ACTCAAAATGGCGCAGATGCCAATTTAGTCTCTTGGAGGTGCAACAGACAGCCCTGGA 2631  
2041 AGCAGTGAAGCTGTGGAAGGGTTTACAAACAGAGCTGTCTCATGATCTCTGATGAC 2100  
2632 GCACAGAGGCTGTGGGCTCTGGCGACAGGCAAGAGCCATTAATTTGTCGCAAGCAAGAT 2691  
2101 AACCGGTACAGAAATGATGGAATCCTCAAGCAGACAAAACGCAATGCTCTCTCTTGTG 2160  
2692 GGAGGACAGCTGGAATCCATGAGTGCCTACAGTATGACAGGCCCTGTGCGAGCCCTTACC 2751  
2161 CAAGAAATGCACAGTCCCATGTGTGAGAGAGCTGCACCTTCACTGCTTGGTCCAAAGTTTACG 2220

2752 CAGGCTGCGCAGATCCCTCGCAGGATGACTGTCAATTGACGAGCTGCTCAAGTTTCT 2811  
2221 CCCTGTCCACGAACTGTGAAGCCACAAAAAGTAGGCGGCGACAGCTCAAGGAAAGC 2280  
2812 TCATGCAATGAGACTGTGTTGAGTGTAGCACCAGAAAGCGCACTCTTGTGGAAAAAGT 2871  
2281 AGAAGAGGAGAAATCCAGGATTTGACCTTTACCTCTAGTGGAGACAGAACTATGT 2340  
2872 AAAAAAATAAATAAATAAATAAATTTCCATTTGATCCCTGATTGAGACTCAGTATTGT 2931  
2341 CCTTGTGATGAATTTATATCCCAACTTATGGAACCTGTGACATTTGATCTTCCAGAA 2400  
2932 CCTTGTGACAAATAATAATGCAACACTGTGGGAACTGTGTCAGACTGTATTTTACCAGAG 2991  
2401 GGCAGAGGAGGAGCCTCACCGAGGACTGCGGGTCAAGCAGACAGCAAGAAATGTGGAGAA 2460  
2992 GGAAGATGGAAGTGTGTGCGGAATGAAGATCAAGGAGACATCAAGGAATGCGGACAA 3051  
2461 GGCCTGGCTTTTCAGCAGTAGCCTGTTCTGATAAAAAATGGAAGACCTGTGACCCCTCC 2520  
3052 GGATATCGTTACCAGCAATGGCATGCTACGATCAAAATGCGAGGCTTGTGGAACAATCT 3111  
2521 TTCTGACGAGCTCTGTTACATTCAGAAAAATGTGCTATTCCCTGCCCTTGTGATTC 2580  
3112 AGATGTAACAGCCATGTTTACATTTGAGGAGCCCTGTCATCATCCCTGCCCTCAGACTGC 3171  
2581 AAGTTAAGCAGATTGGTCTAGTTGGGGTCTTGGAGTTCATCTTGTGGAATTTGGAGTGA 2640  
3172 AAGCTCAGTGAGTGGTCCAACTGCTCGCGCTGCAGCAAGTCTCTGTGGAGTGTGTGAAG 3231  
2641 ATTGATCCAAATGGCTTAAAGAAAAACCTTACAAATGAGAGACAGCAATGTCCCAAACTG 2700  
3232 GTTCGTTCTAAATGGCTGCGTGAAGAAACCATATAATGGAAGGAGCCCTTGCCTCCAACTG 3291  
2701 GATC-----TCAAGAAATCAGGTACATGAGGAGTCCCATGTTTACAGTGTGCAATCAG 2754  
3292 GACCATGTCAACAGGCGACAGGTGTATGAGTGTGCCATGCCACAGTGTGACTGCACAG 3351  
2755 TATTCCTGGTGTGAGAACACTGGTCTTTCATGCAAAATCAACAAATGAGCTGAGTCCCTG 2814  
3352 TACCTATGGTTCACAGAGCCCTGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAG 3411  
2815 CGCTGTGGAGGAGAACACAACTTAGGAAATCAGATGTCT---GAATCTGCGGATGTT 2871  
3412 AACTGTGGAGGCGGTGCAACCCGAAAGAGTGAATGATGCAATGAGAAATACAGCAGATGGC 3471  
2872 GAAGGTGGAGCAGTGGATAGCAACCTGTGCAACCCAGGATGAAATTTCCCGAGAAACCCAG 2931  
3472 CCTTCTGAACATGTAGAGGATTACCTCTGTGACCCAGAGAGATGCCCTGGGCTCTAGA 3531  
2932 TCCTGTTCTTATGTGTCCTCAATGAGTGTGCTATGCTGAGTGGGACTTTTGGAGCAAA 2991  
3532 GTGTGCAAAATACCATGCCCTGAGGACTGTGTATATCTGAATGGGCTCCATGAGCCCAA 3591  
2992 TGCCCAACATGTCAGATCCCAACAAT---GCAAGAAAGAACTCGCCACCTGCTAAGA 3048  
3592 TGTGTTTGGCTTGCATCAAGCAGTTTCGGCAAGGTTCAGCTGATCCCATCAGACAA 3651  
3049 CCATCACTGAACTCAAGACTTGTGTGAAGATCTCAGGTGAGCCCTTGCCTCTGGAAT 3108  
3652 CCAGCTGATGAAGAAAGATCTTGGCTTAATGCTTTGAGAAAGAAACCTGTAAACCTGAAC 3711  
3109 GAAATTTGCTTCCAGTTCCAGTACAATCTAACAGGTGGAGCAGCATCCAGCTGAGTGA 3168  
3712 AAAAACTGTACACTATGATTATATGTAACAGACTGGAGTACATGTCACTGAGTGA 3771  
3169 AACGACCCCTGTGTGAAGCGCTCAGACCCGCTGCTTAAGCTGTGTGTCAGTGTGTC 3228  
3772 AAGCAGTTTGTGGAATGGAATAAAAAAAGAGATGTTGGATGTTGTTTGAAGTGTGTC 3831  
3229 AAGCCAGTCAAGTGAACCAATGTGACAGCATAAATTTGGAGAGCCCAAGAGAAATGAGC 3288



Db 3832 AAGTCAGTGCCTGAAATATTTGTGAAGCGCTGGCTTGGAGAAGAACTGGCAGATGAAC 3891  
Qy 3289 ATTCCTCTGCTGGTGGAAATGGTGTCAACTGTGACGCTCTCAGGTGGACGGCTTCGACA 3348  
Db 3892 ACGTCTTGCAATGGTGGAAATGGCTGTGAATGTGACAGCTTCTGATGGTCTCTTGCTCA 3951  
Qy 3349 GAGTGTTCACAGACTGTGGCCATGGAGGTGGAATGAGACCGGACTCGATTTATCATATG 3408  
Db 3952 GAATGTTCTCAAAATGCTGGCTCACAGGAAATATGATCCGAAGCAACAGTGAACCCAG 4011  
Qy 3409 CCAACCCAAAGGAGAGGACCGCATGCCACAGAGCTTACCAGGAGAAACCTGCCCA 3468  
Db 4012 CCTTTCAAGTGTAGAGACCAATGCCCTTCTCCTGATGACAGCTCCAAACCTGCCCA 4071  
Qy 3469 GTGACCCCTCTCAGCTGGTCTTGGCACTGGTCTGCATGTAAATTTGGAGGTGGA 3528  
Db 4072 GTGAAGCTTGTATCGGTGCAATATGGCCAGTGGTCTCCATGCCAAGTGCAGGAGGCC 4131  
Qy 3529 GACTGTGGGAGAGGATTCAGATCCGAGCGCTTCTGCGATGTGTCACAGTGTCAATA 3588  
Db 4132 CAGTGTGGAGAGGAGCCAGAAACAGGAACATTTCTTGTGTAGTAAGTGAAGGTGAGCT 4191  
Qy 3589 TCTCATGCACTGCAAGCTGTGAGAGTGCATGTGTGAGAAATGCCCTT----- 3638  
Db 4192 GATGATTCACAAAGTGGTGAATGAGGAATTTCTGTGCTGACATTTGAATCATATAGAT 4251  
Qy 3639 -----TGGGACAGATCCTGAAGCAGTGTGTTCTGTGCTTGCCTGGCCAGGAGACTGCCAT 3693  
Db 4252 GGTAAATAAATATGTTCTGGAGAAATCTCGACGCCAGCTTGGCCAGGTGACTGTAT 4311  
Qy 3694 TTAACAGAAATGGTCAGAGTGAGACACATGTAAATTAACCTGCATGTGAAGAAAGCTTT 3753  
Db 4312 TTGAAGACTGGTCTTCTGGAGCTGTGTGACGTGACCTGTGTGATGTGAGGATCTA 4371  
Qy 3754 GAGACTGTGGCCGCCAGTTPAGATCAAGACTTTTATAATTCAGTCTTTTGAAGAACCA 3813  
Db 4372 GGCTTTGGTGAATACAGTTCAGATCCAGACCGGTGATTAACAAGAACTAGAGAATCAG 4431  
Qy 3814 GACAGCTGCCCCCAACAGTTCAGAAACAGCGCTTGTACAGAGGCAATGTTATCAC 3873  
Db 4432 CATCTGTGCCAGACGATGTTAGAAACAAATCATGTTATGATGACAGTGTGATGAA 4491  
Qy 3874 TACACATGGAAGCAAGTCTTTGGAACAAATAACGAACAACTGTATGGTCCAGCGTCA 3933  
Db 4492 TATAATGGAATGGCCAGTGTGGAAGGGCTTCTCCGAAACAGTGTGTTCAAGGTCA 4551  
Qy 3934 GATGGCGTATATTCACAGGAGCTGCTCCCTCAGGCCGCTCCTGCGCATTCGCGAG 3993  
Db 4552 GATGGTATAAATGTAAACAGGGGGTGTGTGTGATGAGCCAGCCCTGTATGCCAGAGTCT 4611  
Qy 3994 TGCAATTCAGCTGCAGAAACCTTTCTCTCTACTGTATACAGAGGTGAGTCTGTGTTGT 4053  
Db 4612 TGTAACCCACGCTGTAGTCAACCCCACTCGTACTGTAGCGAGACAAACATGCCATGT 4671  
Qy 4054 GAGAAGGCTATACAGAGATTAATGAATCAATGGTTCCTGGATTAAGTCA----- 4105  
Db 4672 GAAGAAGGTTCACTGAAGTATGTTCTTAAACAGCACCTTTGAGCAATGCACATTAATC 4731  
Qy 4106 -----TGAAGTACCAAGGTTCAGAGGATAAAGAGTGAATAAAGCTGTAAGAACTTTCTG-G 4157  
Db 4732 CCCGTGGTGTATTAACCAACCATGGAGGACAAAGAGAGATGTGAACACGATCGGGCT 4791  
Qy 4158 GAAAAACAGACTGTGAATCAAAAAATACATGATATTTTAAAGGA--TGGTCTCTTCAA 4215  
Db 4792 GTACATCAACCCCAACCTCCAGTAACCCAGCAGGACGGGAGGAGCACTGGTTCTACAG 4851  
Qy 4216 CCACCTGTATCCAGATGGCCGAGTAAATTTGGTGTATGCGTTTCAGGTGGCGTTTT 4275  
Db 4852 CCAATTTGGCCAGATGGGAGACTAAAGACCTTGGTGTACGGTGTAGCAGCTGGGGCATTT 4911  
Qy 4276 CTCATCATGATTTTCTTAATATTTACTTCTACCTACCTGTGTTTGAAGAGCA--AAACCA 4332  
Db 4912 GTGTTACTCATCTTTATTTGTTCTCCATGATTTATCTAGCTTGAACAAAGCAAGAACCC 4971

Qy 4333 CATCAAGACACACCTCCCAACAGAGCCTCTGACCTTAGCTAGCTAGGAGACTTAGAC 4392  
Db 4972 CAAGAAGGCAAAACACCGACTGAACCTTTAACCCTTAGCTATGATGAGATGCCGAC 5031  
Qy 4393 ATGTAA 4398  
Db 5032 ATGTAA 5037

RESULT 8  
US-10-104-047-958  
; Sequence 958, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 958  
; LENGTH: 3053  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-958

Query Match 13.2%; Score 580.6; DB 15; Length 3053;  
Best Local Similarity 56.4%; Pred. No. 6.8e-178;  
Matches 1309; Conservative 0; Mismatches 934; Indels 78; Gaps 9;

Qy 10 TCATTCAGAGCTGAACCGAAGTGTGTTGCAATGAATATGCGAACAATTTGCCCTT 69  
Db 730 TGCATCCAGAAAGACAAAGACATTCCTCGGAGGATATCATCTGTGAGTACTTTGAGGCC 789  
Qy 70 CAGCCTCTACAGACAGCTTGCCTCATTCCTTGTCCCGGATTTGTAGTATCTCAG 129  
Db 790 AAGCCTCTCTGGAGCAGCTTGCCTCATTCCTTGTCCAGCAAGATTGCATCGTCTGAA 849  
Qy 130 TTCTTACCATGGTCCAACTGTAGCAAGGATGTGGGAAGAAATTCAGCATAGAACTGC 189  
Db 850 TTTTCTGCTGTGTCGAATGTCCAAACCTCGCGCAGCGGCTCCAGCACCGACGGT 909  
Qy 190 GCGGTATAGTCTCCCTCTCTTTGGTGGTTGCAATGTCCAAATCTGACTGACTCAAGA 249  
Db 910 CATGTGTGGCGCTTCCGAGTTCGAGGCTCTGGCTGTCCAAACCTGACGAGTTCCAG 969  
Qy 250 GCCTGTGATGCTCCCATTTCTCTCTCTTGGGGAAGAGGAATATACATTTTAGCCTTAAG 309  
Db 970 GTGTG-----CCAATCCAGTCCATGCGAGCGGAGGAGCTCAGTACAGCTGCAT 1020  
Qy 310 GTTGACCATGGAGTAATGACAGTGCCTCATCTTAAAGAAATTAATCCAGCGGAGA 369  
Db 1021 GTGGGCGCTTGGAGCACCTGTCTCAATGCCCACTCCGCAAGTAAGCAAGCAAGGAGA 1080  
Qy 370 ACTGTTCTGATTTAA-----CTCTGATTTCAATGAG 402  
Db 1081 CCGGGAAGATTAAGACGGGAAGGACCGCAGCAAGGAGTAAGGATCCAGAGCC 1140  
Qy 403 CAGGTCACTTTAAACATCAAGTTTACA-----AAGCACATCATCATTCGAGTCT 453  
Db 1141 CCGAGCTTATTTAGAAAAAGAGAAACAGAAACAGGCAAGCAAGAGACAAATAT 1200  
Qy 454 TGGGCAATAGAGATAGTTATCAACCCGCGAGGTTTCGTGTACAAAGATGATGACAA 513  
Db 1201 TGGGATCCAGATTTGATATCAGACCAAGAGGTTATGTGCAATTAACAGACGGGAAA 1260  
Qy 514 AATGCTATTTAAGCCTTTGGCTTCAAGATTCTTTCCCAATTCAGTGTTCAGTCTGATC 573  
Db 1261 GCTGTGATTTAAGCTTTTGGCAGCAAGAGAGCTTCCAATGACCTTCCAGTCTCTGTG 1320

1624	QY	AAATGCAACCAATTGGCTGGAAATGGAGAGCCACGTGTGGTGTGTAGGCATTCAGACTCGAGA	1683
2401	Db	AACACAACTACGACTTGGAAATGGGGAGGCGCTCTGCTGTGTGGCATGAGACAAGAAAA	2460
1684	QY	GTCTTCTGTGTCAAGAGTCACTGGGACAAAGTAAATGACCAAAAGATGTCCAGATTTCTACT	1743
2461	Db	GTCACTGTGTGCGAGTCAATGTGGGCCAAGTGGGACCAAAAAATGTCTGAAAGCCTT	2520
1744	QY	CGACCTGAAACTGTGGCGCCCTGTTTCTTCCCATGCCAAAAAAGACTGTAATGTGACTGCT	1803
2521	Db	CGACCTGAAACTGTGAAGCGCTGTGTCTGCTTCTTGTTPAAGAGGACTGTATTTGTGACCCCA	2580
1804	QY	TTCACTGAGTGGACACCCCTGCCCAAGATGTGCCAAGCAGGAAATGCCACAGT---	1860
2581	Db	TATAGTGACTGGACATCATGCCCTCTTCGTGTAAAGAGGGGACTCCAGATATCAGGAAG	2640
1861	QY	CAGTCTCGATACAGAATCATATCCAAAGACAGGCCAAATGGAGGCCAGGAATGCCCCAGAT	1920
2641	Db	CAGTCTAGGCATCGGGTCAATCATTCAGTCCGACCAACGGGGCCGAGACTGCACAGAT	2700
1921	QY	ACCTTATATGAGGAGAGAGTGTGAAGATGTTTTCTTGTGTCTGTATATCGGTGGGAAG	1980
2701	Db	CCCTCTATGAAGAGAGGCCCTGTGAGGCACCTCAAGCGTGCCTCAAAAGCTACACGTGGGAAG	2760
1981	QY	CCACAAATGGAGCCCTTGCACTCTTAGTGCCACAGACTGTCTGGCAGGGAATAACGGGC	2040
2761	Db	ACTCAAAATGGCGCAGATGCCAAATTAGTCCCTTGGAGCGTGCACAAGACAGCCCTGGA	2820
2041	QY	AGCAGTGAAGCCTGTGGAAAGGGGTTACAAAACAGAGCTGTCTCATGCATCTCTGATGAC	2100
2821	Db	GCACAGGAAGGCTGTGGGCGCTGGGGCAGCAGGCAAGAGCCATTACTTGTGCGAAGACAGAT	2880
2101	QY	AACCGGTCAAGCAAAATGATGGAATGCCTCAAGCAGACAAAACGGCATGCCTCTCTTGTG	2160
2881	Db	GGAGGACAGCTGGAATCCATGAGTGCCTACAGTATGTCAGGCGCCTGTGCCAGGCCCTTACC	2940
2161	QY	CRAGAATGCACAGTCCCATGTGGAGAGACTGCACCTTCACCTGCTTGGTCCCAAGTTTACG	2220
2941	Db	CAGGCTGCCAGATCCCTGCCAGATGACTGTCAATTGACACAGCTGGTCCCAAGTTTCT	3000
2221	QY	CCCTGTCCACGAACCTGTGAAGCCACAAAAAGTGGCGCG	2261
3001	Db	TCATGCAATCGAGACTGTGGTGCAGTTAGGACACAGAAAGCG	3041

RESULT 9  
US-09-960-253-20

	Query Match	11.7%	Score 512.4	DB 9	Length 516
	Best Local Similarity 99.8%	Pred. No. 3.9e-156			
	Matches 513	Conservative 0	Mismatches 1	Indels 0	Gaps 0
Qy	1713	AGTAATGACCAAAAGATGTCAGATTCTACTCGACCTGAAACTGTGCGCCCTGTTTCT	1772		
Db	3	AGAAATGACCAAAAGATGTCAGATTCTACTCGACCTGAAACTGTGCGCCCTGTTTCT	62		

1773 CCATGTCAGAAAGACTGTATTGTGACTGCTTTTCACTGAGTGGACACCCCTGCCACAGGAT 1832  
Db 63 CCCATGTCAGAAAGACTGTATTGTGACTGCTTTTCACTGAGTGGACACCCCTGCCACAGGAT 122  
1833 GTGCCAAGCAGGAAATCCACAGTAAACAGTCTCGATACAGAAATCATCATCCAAAGAGC 1892  
Db 123 GTGCCAAGCAGGAAATCCACAGTAAACAGTCTCGATACAGAAATCATCATCCAAAGAGC 182  
1893 AGCCAAATGGAGGCCAGAAATCCAGATACCTTATATGAGGAGAGAGTGTGAAGATGT 1952  
Db 183 AGCCAAATGGAGGCCAGAAATCCAGATACCTTATATGAGGAGAGAGTGTGAAGATGT 242  
1953 TTCCTTGTGCTCTATATCGTGGAAAGCACAGAAATGAGCCCTTGCACTCTTAGTGCC 2012  
Db 243 TTCCTTGTGCTCTATATCGTGGAAAGCACAGAAATGAGCCCTTGCACTCTTAGTGCC 302  
2013 AGAGTCTGTCTGGCAGGAAATACCGGCAGCAGTGAAGCCCTGTGGAAAGGGTTACAAC 2072  
Db 303 AGAGTCTGTCTGGCAGGAAATACCGGCAGCAGTGAAGCCCTGTGGAAAGGGTTACAAC 362  
2073 AAGAGCTGTCTCATGATCTCTGATGACAAACCGGTGAGCAAGAAATGATGAATGCTCAA 2132  
Db 363 AAGAGCTGTCTCATGATCTCTGATGACAAACCGGTGAGCAAGAAATGATGAATGCTCAA 422  
2133 GCAGACAAACGGCAGTCTCTCTTGTGCAAGAAATGACAGTCCCATGTGCGAAGACTG 2192  
Db 423 GCAGACAAACGGCAGTCTCTCTTGTGCAAGAAATGACAGTCCCATGTGCGAAGACTG 482  
2193 CACCTTCACTGCTTGGTCCAGTTTACGCCCTGC 2226  
Db 483 CACCTTCACTGCTTGGTCCAGTTTACGCCCTGC 516

RESULT 10

US-09-864-761-27172  
; Sequence 27172, Application US/09864761  
; Patent No. US2002048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263,6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 27172  
; LENGTH: 744  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC011231.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.89  
; OTHER INFORMATION: EST HUMAN HIT: BE295719.1, EVALUE 6.00e-53  
; OTHER INFORMATION: NT HIT: AF133643.1, EVALUE 1.90e+00  
; OTHER INFORMATION: SWISSPROT HIT: P35446, EVALUE 6.00e-10  
US-09-864-761-27172

Query Match 10.5%; Score 460; DB 9; Length 744;

Best Local Similarity 100.0%; Pred. No. 7.4e-139;

Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGCGCTGCATTCAGAGCTGAACCGAGCTGCTGCTGCAATGAATATGCGAACAC 60  
Db 285 ATGTGCGCTGCATTCAGAGCTGAACCGAGCTGCTGCTGCAATGAATATGCGAACAC 344  
QY 61 TTTTGCCCTTCAGCCTCTCTACAGAACAGAGCTTGCCTCATCTCTCCCGGGATTGTGA 120  
Db 345 TTTTGCCCTTCAGCCTCTCTACAGAACAGAGCTTGCCTCATCTCTCCCGGGATTGTGA 404  
QY 121 GTATCTGAGTCTTACCATGCTCCAACTGACAGAGGATGCGGAGAAATTCAGCAT 180  
Db 405 GTATCTGAGTCTTACCATGCTCCAACTGACAGAGGATGCGGAGAAATTCAGCAT 464  
QY 181 AGAATCTCGCGGGTGCATAGCTCCCGCTCTCTTTGGTGGTTTGCATGTCCAAATCTGACT 240  
Db 465 AGAATCTCGCGGGTGCATAGCTCCCGCTCTCTTTGGTGGTTTGCATGTCCAAATCTGACT 524  
QY 241 GAGTCAAGAGCTGTGATGCTCCCATTTCCCTCTCTTTGGGAGAGAGATATACATTT 300  
Db 525 GAGTCAAGAGCTGTGATGCTCCCATTTCCCTCTCTTTGGGAGAGAGATATACATTT 584  
QY 301 AGCCTTAAGGTTGGACCATGGAGTAATGAGAGTGCCTCATCTTAAAGAAATTAATCCA 360  
Db 585 AGCCTTAAGGTTGGACCATGGAGTAATGAGAGTGCCTCATCTTAAAGAAATTAATCCA 644  
QY 361 AGCGGAAGAACTGTTCTGGATTTTAACTTCTCAATCAATGAGCGAGTCACTTTAAACAT 420  
Db 645 AGCGGAAGAACTGTTCTGGATTTTAACTTCTCAATCAATGAGCGAGTCACTTTAAACAT 704  
QY 421 CAAAGTTTAAAGCAATCATCATTTTGAAGTCTTGGGCAA 460  
Db 705 CAAAGTTTAAAGCAATCATCATTTTGAAGTCTTGGGCAA 744

RESULT 11

US-10-029-386-5068/c  
; Sequence 5068, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS TWO

RESULT 14

US-10-029-386-22453/c

; Sequence 22453, Application US/10029386

; Publication No. US20030194704A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G

; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

; FILE REFERENCE: AECOMICA-X-2

; CURRENT APPLICATION NUMBER: US/10/029,386

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 34288

; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1

; SEQ ID NO 22453

; LENGTH: 174

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AC012000.3

; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1

; OTHER INFORMATION: EXPRESSED IN LIVER, SIGNAL = 0.53

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.81

; OTHER INFORMATION: NT HIT: g14721097, EVALUATE 2.00e-93

; OTHER INFORMATION: SWISSPROT HIT: P36297, EVALUATE 1.10e+00

; OTHER INFORMATION: EST\_HUMAN HIT: BE940082.1, EVALUATE 8.90e-01

US-10-029-386-22453

Query Match

Best Local Similarity 4.0%; Score 174; DB 14; Length 174;

Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1103 GATTAGAACGAGCAGCGCCATGCTCTCATGGAATCTACAGGCGCTGCAGGCGCATGGCC 1162

DB 174 GATTAGAACGAGCAGCGCCATGCTCTCATGGAATCTACAGGCGCTGCAGGCGCATGGCC 115

QY 1163 CTCATTGGTGAGCTGTTCTTCTGTGAGGATCCAAATGCTACCGATGCTGCGCATCAG 1222

DB 114 CTCATTGGTGAGCTGTTCTTCTGTGAGGATCCAAATGCTACCGATGCTGCGCATCAG 55

QY 1223 AAGGATCTGTTCCCTGATCATGGAATAATGTGGCCCTGGGACATCGTATTCTGA 1276

DB 54 AAGGATCTGTTCCCTGATCATGGAATAATGTGGCCCTGGGACATCGTATTCTGA 1

RESULT 15

US-09-919-172-96

; Sequence 96, Application US/09919172

; Patent No. US20020119463A1

; GENERAL INFORMATION:

; APPLICANT: Paris, Mary

; APPLICANT: Turner, Christopher M.

; TITLE OF INVENTION: PROSTATE CANCER MARKERS

; FILE REFERENCE: PA-0036 US

; CURRENT APPLICATION NUMBER: US/09/919,172

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: 60/222,469

; PRIOR FILING DATE: 2000-07-28

; NUMBER OF SEQ ID NOS: 102

; SOFTWARE: PERL Program

; SEQ ID NO 96

; LENGTH: 579

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: misc feature

; FEATURE:

; OTHER INFORMATION: Incyte ID No. US20020119463A1 041764.1

US-09-919-172-96

Query Match

Best Local Similarity 3.7%; Score 161.8; DB 9; Length 579;

Matches 340; Conservative 0; Mismatches 227; Indels 12; Gaps 3;

QY 2662 GAAAAACCTTACATGAGGACGACCACTGTCCTCCAACTGGATCTCAAGATCAGGTAC-- 2719

DB 1 GAAAAACCATATAATGAGGAGAGGCCCTTGCCCCAACTGGACCAATGTCAACACGACAG 60

QY 2720 ----ATGAGGAGTCCCATGTTTACAGTGAAGTGCATCAGTATTCTCGGTTGTAGAACAC 2775

DB 61 GTGTATGAGGTTGTCCCATGTCACAGTGACTGCAACCCAGTACCTATGGTGCACAGAGCCC 120

QY 2776 TGGTCTTCATGCAAAAATCAACAATGAGTGAAGTCCCTGGCTGTGGAGGAGGACACAA 2835

DB 121 TGGAGCATCTGCAAGGTGACCTTTGTGAATATGCGGAGAACTGTGGAGAGGGCGTGCAA 180

QY 2836 TCTAGGAAAATCAGATGTGT---GAATACTCGGATGCTGAAGGTGGAGCAGTGGATAGC 2892

DB 181 ACCCGAAAAGTGAGATGATGCAGATACAGCAGATGCGCCCTTCTGAACATGTAGAGAT 240

QY 2893 AACCTGTGCAACAGGATGAAATTTCCCCCAGAAAACCCAGTCCCTGTTCTTATGTGTCCT 2952

DB 241 TACCTCTGTGACCCAGAGAGATGCCCCCTGAGAGTGTGCAAAATACCAATGCCCT 300

QY 2953 AATGAGTGTGTCTGTCTGAGTGGGAGCTTTGGAGCAAAATGCCCAAGTATGCGATGCC 3012

DB 301 GAGGACTGTGTGATATCTGAATGGGTCCATGGACCCCAATGTGTTTGCCTTGGCAATCAA 360

QY 3013 CACACAAT---GCGAGAGAAAGTGGCCACCTGCTAAGACCATCACTGAACCTCAAGACT 3069

DB 361 AGCAGTTTCCGGCAAAAGGTCAAGTCTATCCCATCAGACAAACCCAGCTGATGAAGGAATCT 420

QY 3070 TGTGCTGAAGACTCACAGGTGCGCCTTGCCCTGCTGAATGAAAATGCTTCCAGTCCAG 3129

DB 421 TGCCCTAATGCTGTTGAGAAAGAACCTGTAACTGAACAAAACCTGCTACCATATGAT 480

QY 3130 TACAATCTAACAGAGTGGAGCACATGCGCAGCTGAGTGAAAACGACCCCTGFGTCAAGGC 3189

DB 481 TATAATGTAACAGACTGGAGTACATGTGAGTGAAGAGGCAAGTTTGTGGAATGGA 540

QY 3190 GTCAGGACCCCTGCTAAGCTGTGTGTCAGTGTGCTG 3228

DB 541 ATAAAAACAAGATGTTGGATTGTTTCAAGTGTATGCG 579

Search completed: February 24, 2004, 00:59:43

Job time : 1490.49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: February 23, 2004, 20:29:46 ; Search time 134.011 Seconds  
(without alignments)  
3088.802 Million cell updates/sec

Title: US-10-022-710-2

Effect score: 8241

Sequence: 1 MVRCTQKLNRTVANEICEH.....OSTPPQKRLTLAYDGDLDL 1465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	8227	99.8	1574	7	Ade08053 Novel pro
2	4296	52.1	1551	4	Aam41081 Human pol
3	4296	52.1	1588	4	Aab20155 Secreted
4	4296	52.1	1588	6	Abr58456 Human NOV
5	4296	52.1	1588	6	Abr58456 Human NOV
6	4296	52.1	1624	5	Aae23979 Human LP2
7	4292	52.1	1588	4	Aam39295 Human pol
8	4290	52.1	1588	6	Abr58450 Human NOV
9	4284	52.0	1588	6	Abr58459 Human NOV
10	4113.5	49.9	1549	6	Abr58453 Human NOV
11	4095.5	49.7	1549	6	Abr58455 Human NOV
12	4033	48.9	1490	6	Abr58454 Human NOV
13	2729	33.1	484	5	Aao15279 Human lar
14	2126	25.8	933	7	Adb64774 Human pro
15	1762.5	21.4	577	6	Abr58460 Human NOV
16	1756.5	21.3	571	6	Abr58452 Human NOV
17	1389.5	16.9	535	6	Abr58458 Human NOV
18	1389.5	16.9	535	6	Abr58457 Human NOV
19	1361.5	16.5	467	6	Abr58451 Human NOV
20	1332.5	16.2	617	3	Aab42496 Human ORF
21	1121.5	13.6	284	4	Abg05063 Novel hum
22	852	10.3	247	4	Aam70792 Human bon
23	852	10.3	247	4	Aam58317 Human bra
24	812	9.9	149	4	Abg05062 Novel hum
25	694	8.4	2150	3	Aay53898 Amino aci

## ALIGNMENTS

### RESULT 1

ADE08053

ID ADE08053 standard; protein; 1574 AA.

XX AC ADE08053;

XX DT 29-JAN-2004 (first entry)

XX DE Novel protein (useful for identifying genetic disorders) #208.

XX KW novel gene; novel protein; tissue marker; molecular weight marker;

XX KW chromosome marker; genetic disorder.

XX OS Unidentified.

XX FN WO2003054152-A2.

XX PD 03-JUL-2003.

XX PF 10-DEC-2002; 2002WO-US039555.

XX PR 10-DEC-2001; 2001US-0339739P.

XX PR 11-DEC-2001; 2001US-0339453P.

XX PR 14-MAR-2002; 2002US-0365091P.

XX PR 14-MAR-2002; 2002US-0365384P.

XX PR 12-APR-2002; 2002US-0372381P.

XX PR 12-APR-2002; 2002US-0372815P.

XX PR 22-APR-2002; 2002US-00128558.

XX PR 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

XX Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;

XX Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI; 2003-569235/53.

XX N-PSDB; ADE07142.

XX DR New polynucleotides, useful for expressing recombinant proteins for

XX PT analysis, characterization or therapeutic use, or as markers for tissues

XX PT in which the corresponding protein is preferentially expressed.

XX PS Claim 20; SEQ ID NO 1119; 1177bp; English.

XX CC The invention comprises the amino acid and coding sequences of novel

XX CC proteins. The DNA and protein sequences of the invention are useful as:

XX CC markers for tissues in which the corresponding protein is preferentially

Aab90617 Human sec  
Aae19173 Human pro  
Aau71133 Human pro  
Aau72896 Human met  
Aab72286 Human ADA  
Abu62079 Human jel  
Aab72301 Human ADA  
Abg30203 Novel hum  
Abg21064 Novel hum  
Abg30703 Human agg  
Aao14448 Human ADA  
Abg30702 Human agg  
Abr58461 Human NOV  
Aab01431 Human TAN  
Aao30834 Human cel  
Aau9587 Novel hum  
Aau80189 Human TSP  
Aab90551 Human sec  
Abg65473 Human alb  
Abb10246 Human CDN

26 694 8.4 2165 4 AAB90617  
27 674 8.2 1916 5 AAE19173  
28 673 8.2 1907 5 AAU71133  
29 673 8.2 1935 5 AAU72896  
30 657 8.0 1882 4 AAB72286  
31 655.5 8.0 4123 7 ABU62079  
32 646 7.8 1934 4 AAB72301  
33 598.5 7.3 4561 4 ABG30203  
34 598.5 7.3 9222 4 ABG21064  
35 597.5 7.3 1629 5 ABG30703  
36 597.5 7.3 1629 5 AAO14448  
37 578.5 7.0 1602 5 ABG30702  
38 547.5 6.6 307 6 ABR58461  
39 507.5 6.2 874 3 AAB01431  
40 498 6.0 1762 7 AAO30834  
41 495 6.0 1762 5 AAU9587  
42 482.5 5.9 966 5 AAU80189  
43 477.5 5.8 1745 4 AAB90551  
44 477.5 5.8 1745 5 ABG65473  
45 477.5 5.8 1766 4 ABB10246

CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present amino acid sequence represents a protein  
CC of the invention.

Query Match	99.8%; Score 8227; DB 7; Length 1574;
Best Local Similarity	99.8%; Pred. No. 0;
Matches 1464; Conservative 1; Mismatches 0; Indels 2; Gaps 1;	
QY	1 MVRCIQKLNRTWANEICEHFALQPPEQACILPCPRDCVSEFLPWNCCKGKGKLOH 60
Db	108 MVRCIQKLNRTWANEICEHFALQPPEQACILPCPRDCVSEFLPWNCCKGKGKLOH 167
QY	61 RTRVIAPIPLFGGLQCENLFTESRACDAPISCPGLBEEYFTSLKVPFWSKRLPHLKEINP 120
Db	168 RTRVIAPIPLFGGLQCENLFTESRACDAPISCPGLBEEYFTSLKVPFWSKRLPHLKEINP 227
QY	121 SORTVLDNFNSDINERVTFKQSYKAHHHSKSWAIEIGYQTRQVSTCRSDGQVAMLSLCLQ 180
Db	228 SORTVLDNFNSDINERVTFKQSYKAHHHSKSWAIEIGYQTRQVSTCRSDGQVAMLSLCLQ 287
QY	181 DSFPLTVQSCIMPDKDCTSONSWSPCKSTRSGSLLPGRFSRSRNKVMKMAIGGKGCPCE 240
Db	288 DSFPLTVQSCIMPDKDCTSONSWSPCKSTRSGSLLPGRFSRSRNKVMKMAIGGKGCPCE 347
QY	241 LLEKACACIVEGELLQCCPRYSWRTSEWKEQVSLLEQQDPHHVTFGPVCGGGIQTREYV 300
Db	348 LLEKACACIVEGELLQCCPRYSWRTSEWKEQVSLLEQQDPHHVTFGPVCGGGIQTREYV 407
QY	301 CAQSVPAALAAKESRPEVKALCVGPAPLPQCLNIPCSSTDCTIVSSWSAWGLCIHENC 360
Db	408 CAQSVPAALAAKESRPEVKALCVGPAPLPQCLNIPCSSTDCTIVSSWSAWGLCIHENC 467
QY	361 HEPQCKGFRTRQHVLMESTGPAGHCPLHVESVPCEDPMCVRWLASGICFPDHGKGL 420
Db	468 HEPQCKGFRTRQHVLMESTGPAGHCPLHVESVPCEDPMCVRWLASGICFPDHGKGL 527
QY	421 GHRILKAVCONDRGEDVSGSLCPVPPPERKSCIEIPCRMDCVLSEWTWSSCSOSCSNKN 480
Db	528 GHRILKAVCONDRGEDVSGSLCPVPPPERKSCIEIPCRMDCVLSEWTWSSCSOSCSNKN 587
QY	481 SDGKOTRSTRTILALAGEGKPCPSQALQHRILCNDSQWLHWETSPWGCSDETLVTA 540
Db	588 SDGKOTRSTRTILALAGEGKPCPSQALQHRILCNDSQWLHWETSPWGCSDETLVTA 647
QY	541 LNATIGWNGEATCGVIGITRRVFCVKSHVQVNTKCPDSTRPETHVPCFLPCKKOCIVT 600
Db	648 LNATIGWNGEATCGVIGITRRVFCVKSHVQVNTKCPDSTRPETHVPCFLPCKKOCIVT 707
QY	601 AFSEWTPCPMCOAGNATVKQSRVLTIIQEAANGQCECDTLTYEERECEDVSLCPVTRWK 660
Db	708 AFSEWTPCPMCOAGNATVKQSRVLTIIQEAANGQCECDTLTYEERECEDVSLCPVTRWK 767
QY	661 POKWSFPCILVPESVMQGITGSSEACGKGLQTRAVSGISDDNRSAMWELCKQTGMPLV 720
Db	768 POKWSFPCILVPESVMQGITGSSEACGKGLQTRAVSGISDDNRSAMWELCKQTGMPLV 827
QY	721 QECITVPCREDCTFTAWSKFTPCSTNCEATKSRRLQTLGSKRKEKQSDIYPLIVETELC 780
Db	828 QECITVPCREDCTFTAWSKFTPCSTNCEATKSRRLQTLGSKRKEKQSDIYPLIVETELC 887
QY	781 PCDEFISQYGNWSDCILPEGRREPHRGLRVQADSKCEGGLRFRVAVACSDKNGRPVDPDS 840
Db	888 PCDEFISQYGNWSDCILPEGRREPHRGLRVQADSKCEGGLRFRVAVACSDKNGRPVDPDS 947
QY	841 FCSSSGYIQEKCVLPCPPDCKLSDWSSWGS CSSCGIGYIRIRSKMLKEKPYNGRPPCKL 900
Db	948 FCSSSGYIQEKCVLPCPPDCKLSDWSSWGS CSSCGIGYIRIRSKMLKEKPYNGRPPCKL 1007
QY	901 DLKN--QVHEAVPYCYSECNOYSWVHEHSSCKTNBELRSIRCGGGTQSRKRCVNTADGE 958



29-NOV-2000; 2000US-00727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
Zhou P, Goodrich R, Drmanac RT;

WPI: 2001-442253/47.

N-PSDB; AA160237.

Novel nucleic acids and polypeptides, useful for treating disorders such  
as central nervous system injuries.

Example 2; SEQ ID NO 6012; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the  
encoded polypeptides (AA158442-AA162213) with nootropic,  
immunosuppressant and cytostatic activity. The polynucleotides are useful  
in gene therapy. A composition containing a polypeptide or polynucleotide  
of the invention may be used to treat diseases of the peripheral nervous  
system, such as peripheral nervous injuries, peripheral neuropathy and  
localised neuropathies and central nervous system diseases, such as  
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
lateral sclerosis, and Sny-Drager Syndrome. Other uses include the  
utilisation of the activities such as: Immune system suppression,  
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
assays for receptor activity, arthritis and inflammation, leukaemias and  
C.N.S disorders. Note: The sequence data for this patent did not form  
part of the printed specification

Sequence 1551 AA;

Query Match 52.1%; Score 4296; DB 4; Length 1551;

Best Local Similarity 50.7%; Pred. No. 2.8e-246;

Matches 766; Conservative 238; Mismatches 449; Indels 59; Gaps 20;

2 VRCIQKNTTVANIECEHFALQPTTEQACLPDRCDVVSFLPWSNCKGCGKKLOHR 61

52 IACIQKNDIPAEADIIICEYFEPEKPLEEQACLPQCDCTVSEFSAMSECKTGGGLQHR 111

62 TRAVIAPPLFGLQCNLTESACDAPISCLPEEYTFSLXGVPWSKRLPHLKEINPS 121

112 TRHVAPPOFGSGCNPNTLTFVQCQ---SSPEAEELRYSLHVGFWSTCSMPHSQVROQA 169

122 GRT-----VLDNFSDSNERTVTK-HQSYKAHHHSKSWAIEIGYQTRQVSCTR 167

169 RRRGKNGKEREKDRSKGVKD--PEARELIKKNRNRRQENKYNWDIQIGYQTRVWMCIN 226

168 SDQGNAMLSLCLQDSPPLTVQSCIMPQDCETQWSWSPSCSKTCBSGSLLPGRFSRSNV 227

227 KTGKAADLSFCQEKLPMTFQSCVIITKEQVSEWSEWSPSCSKTCHDMVSPAGTRVTRTI 286

228 KMAIGGKCEPELLEKEACIVEGELLQCPRYSWRTSEWKECVQSLLEQQDPHHVYG 287

287 RQFFIGSEKEFEFEKEPCLSGDGVWPCATYGNWTTIEWTECRVDPPLSQDKERGNQT 346

288 PIVCGGIQTRVYVCAQS-----VPAALRAKEVSRPVEKALCVGPAPLPSQLCNTPCSTD 343

347 ALCGGGIQTREYVCYQVQANENLLSLSLTHKNKASKPMDLKLCTGPIPTNTQLCHIPCPE 406

344 CIVSSWSAMGLCITHENCHPEQKKGKFRTRQRHVLMESTGPA--GHCPLHVESVCEPDM 400

407 CEVSPHSANGPTCYENCNDQKKGKFLKRRITNEPTGGSGVTGNCPLHLEAIFCEPEA 466

401 CYRWLASE-GICPPDHGK-CGLGHLRLKAVCONDRGEDVSGSLC-----PVPPPPERKSC 453

467 CYDWKAVRLGDCPEPDNGKEGCGPTQVEVVCINSQGEVDRLQDRDAIFPIP-----VAC 521

454 ETPCRMDCVLSWETWESSQSCSNKNSGKQTRSRITLALAG-EGGKPCPSQALQSHR 512

522 DAPCPKDCVLSWSTWSSCSHTCSGKTTGKQIRAKSILAYAGEGGIRCPNSSLQVQR 581

QY 513 LCNDSHSCMQLHWETSFWGPCSEDTLVTALNATIGWNGEATCGVIGIOTRRVFCVKSHVGV 572  
Db 582 SCNEHPTCVVHWOTGPWGQIEDTSVSSFNITTTWNGEASCSVMQTRKVICRVNVGV 641  
QY 573 MTKRCPSDTEPEVTRPCFLPCKKXDCITVAFSEWTPCPRMCOAGNATV-KOSRYRIIIEA 631  
Db 642 GPKKCPESLAPETVRPCLLPCKKDCIVTFSWTSVSCSSCKEGSSIRKQSRHVIQLP 701  
QY 632 ANGOQCEPDTLYEERCEDEVSCLPVWRPQKWSPCILVPESVWQITGSSSEACGKGLQT 691  
Db 702 ANGRDCTDLYEKEACEAPQACQSYRWKTHKWHRCQLVPSVQSDSPGAQEGCGPQRA 761  
QY 692 RAVSCISDDNRSAAEMMECLKQTNMGMLLVQECTVPCKEDCTFTAMSKFTPCSTNCEATKS 751  
Db 762 RAITCRKQDQAGIHECLQYAGVPALTAACOIPQDDCOLTSWKSFCSCNGCCAVRT 821  
QY 752 RRQLTKSKKKEKQSDSLYPLVETELCPDFFISQPYGNWSDCILPBGRRPHGLRV 811  
Db 822 RKRTLVSKEKKECKNSHLYPLIETQYCPDKYNAQPVGNWSDCILPBGKVBVLLGMKV 881  
QY 812 QADSKEGGLRFRVACSDKNGRPVDPSPCSSGYIOEKVIPCPDCKLSDWSSWGSC 871  
Db 882 QGDIKEGCGGIRIQAWACYDQNGRLVETSRNSHGHIIEACIIPCFSDCKLSWNSRC 941  
QY 872 SSSCGIGVIRIRSKWLKEKPYNGRCPCKLDLKN--QVHEAVPCYSECNQYSWYVEHSSC 929  
Db 942 SKSCGSGKVRSKWLREKPYNGRCPCKLDHVNQAQVYVWVCHSDCNQYLVWTEPWSIC 1001  
QY 930 KINNELSLRCGGTQSRKIRCV-NTADGEGGAVDSNLQNDIPIPETOSCSLMCPNECV 998  
Db 1002 KVTFFVNNRENGEGVQTRKVRKQNTADGSEHVEDYLCDEEMPLGSRVKLPCBEDCV 1061  
QY 989 MSEWGLWSKCPQSCDPHTMORTREHLIR-PSLNSRTCAEDSQVQPCLLNENCFQFQYNTL 1047  
Db 1062 ISEWGPWTQCVLPQCNQSSPQRSADPIQPADSEGRPCNAVEKEPCNLKNKYHYDNTV 1121  
QY 1048 EWSTCQLSENAQCGQVTRRLLSVCVSDGKPVSMDOCEOHNLKPKORMSIPCLVECVVNC 1107  
Db 1122 DWSTCQLSEKAVCGNGIKRMLDCVRSKGSDVLDKYCEALGLEKNQMNTSCMVECPVNC 1181  
QY 1108 QLSGWTAWTSCSTCGHGGRMSRTRFIIMPTQCEGRPCPTLTQETCTPVTPCYSVVLGN 1167  
Db 1182 QLSDWSPWSECSQTCGLTGKMIERRVTQPFQDGRPCPSLMDQSKPCPKYRQYQG 1241  
QY 1168 WSACKLEGCGCGVQIRLSLSCMVHSGSISHAAGRYEDALCGEMPF-----QDSILKQLC 1222  
Db 1242 WSPCQVQEAQCGEGTRTNISCVVSDGSDADDFSKVVDDEEFCADIELIIDGNKNWLEESC 1301  
QY 1223 SVPCPGDCHLTSEWSEWSTCELTCDIGRSFETVGRQSRRTFIIOSEFENODSCPOQVLETR 1282  
Db 1302 SQPCPGCYLKMSSWSLCLQTCVNGEDLGFQIGVRSRPIVIOELENOHLCEQMLETR 1361  
QY 1283 PCTGGKCYHTWKAASLWNNNERTWQCRSDGVNVTGCSFQAPAAIROCIIPACRPFYS 1342  
Db 1362 SCYDGCQCYEYKMAWASAKGSSRTVWCQSDGINVTGCLVMSQPDADRSNPPCSPHSY 1421  
QY 1343 CTQGGVCGCKGTYETKNSNGFLDYC-----MKVPSSEDKADYKLSGKRPVNSKIHD 1397  
Db 1422 CSETKTCHCEGTEVMSNSNSTLEQCLTIPVVVLPMTMEDKRGDKVT-SRAVHPTQSSNP 1480  
QY 1398 IFXG--WSLOPLPDGRVKIWWYGVGGAFILIFLIFTSYLVCKKP-KPHQSTPPQOKP 1454  
Db 1481 AGRGRTWFLQPPFDGELKTTWYGVGAAGAVLLIFIVSMYILACKKPKPQRRQNNLKP 1540  
QY 1455 LTLAYDGDLDL 1465  
Db 1541 LTLAYDGDADM 1551

RESULT 3

AAH20155

ID AAB20155 standard; protein; 1598 AA.







```

1048 EWSTCOLSENAPCGQGVTRILLSCVSCDGPYSMDQCEQHNLEKQPMRSIPCLVECVNVC 1107
1159 DWSTCOLSEKAVCGNGIKTRMLDCVRSDDGSKVDLKYCEALGLEKNQWMTSCWVECPVNC 1218
1108 QLSGWTAWTECSOTCGHGRMSRTRIIIMPTQOEGRPPCTELTOEKTCPVTTCYSWVLGN 1167
1219 QLSDWSPWSECSOTCGLTGKMRRTVTQTFQDGRPCPSLMDQSKPCVPKPCYRMYGQ 1278
1168 WSAKLEGGCGGVOIRSLSCMVHSGSISHAAGRVEDALCGEMPF-----ODSIILKQLC 1222
1279 WSPCVQEAQCGEGTTRNISCVVSGSADDDPSKVVDDEFCADIELIIDGNKNWLEESC 1338
1223 SVPCPDGCHLTWEMSESTCELTICIDGRSPETVGRQSRRTFIIQSPENQDSCPQVLETR 1282
1339 SQPCPDGCVLKDWSHSLCQLTCVNGEDLFGGIGQVRSRPVITQELNHLCPQOMLETK 1398
1283 PCTGGKCYHYTWKASLWNNERTVWCQRSDGVNVTGCGSPQAPPAIROCIIPACRPFESY 1342
1399 SCVDGQCYEYKNWASAKGSSRTVWCQRSDGINVTGCGLVMSQPDADRSCNPPCSQPHSY 1458
1343 CTQGGVCGCEKGYTEINKNGFLDYC-----MKVPGSEDKADVKNLGKNRPVNSKIHD 1397
1459 CSETKTCHCEGTEVWSSNSTLEQCTLIPVVVLPTWEDKGDVKT-SRAVHPTQSSNP 1517
1398 IFKG--WSLOPLDPDGRVKIWWYGVSGGAFIMFLIFTSYLVCKKP-KPKQSTPPQOKP 1454
1518 AGRGRWFLQPFQFPGDGRLLKTVYVGAAGAFVILIFTVSMIYLACKPKPKPQRQNNRLKP 1577
1455 LTLAYDGDLDLM 1465
1578 LTLAYDGDADM 1588

RESULT 6
AAE23979
ID AAE23979 standard; protein; 1624 AA.
XX AAE23979;
AC AAE23979;
DT 23-SEP-2002 (first entry)
XX Human LP217 secreted protein.
DE Human; secreted protein; atherosclerosis; Alzheimer's disease; LP217;
KW diabetic retinopathy; severe combined immunodeficiency; pancreatitis;
KW rheumatoid arthritis; colorectal adenoma; haemolytic anaemia; cancer;
KW reperfusion injury; arteriosclerosis; wound healing; transgenic animal;
KW gene therapy; neoplasm; transgenic; psoriasis; ischaemia; carcinoma;
KW chromosome 7p21-p22.
XX Homo sapiens.
OS Homo sapiens.
PH Key Location/Qualifiers
FT Peptide 1..115
FT Protein /label= signal_peptide
FT 16..1624
FT /note= "Mature human LP217 secreted protein"
XX W0200226801-A2.
PD 04-APR-2002.
XX 14-SEP-2001; 2001WO-US026026.
XX 28-SEP-2000; 2000US-0236088P.
XX (ELIL ) LILLY & CO ELI.
XX Su EW, Wang H;
XX WPI; 2002-471259/50.
XX N-PSDB; AAD38694.
XX

```

```

2 VRCIQKLNRTVAVNICEHFALQPTEQACILPCPRDCVWSEFLPWSKSCGCGKLOHR 61
89 IACIQKDKDIPAEIDICEYFEKPLEQACILPCQODCIVSEFSANSECSKTCGSGLOHR 148
62 TRAVTAPPLFGGLQCPNLTESACDAPISCPGEBEYFSLKVGWPSKCRILPHUKENIPS 121
149 TRHVAVPQFGSGCPNLTEFQVCQ---SSPCEABELRYSLHVGFWSTCSPHSPRQVQA 205
122 GRT-----VLDNFSDSNERTVEFK-HOSYKAHHHSKSWAIBIGYCTQVSCVCTR 167
206 RRRGNKREKDRSGKVD--PEARLTKKXNRNRQNRQENKYWDIQIGYQTRVWMCIN 263
168 SQQNAMLSLCLODSFPLTVQSCINPKXCETSQWSSWSPCKTKCRSGSLLPFGFRSRNV 227
264 KTGKAADLSFCQOEKLPMTFQSCVITKEQVSEWSEWSPCKTKCHDMVSPAGTRVTRTI 323
228 KMAVJGGGKECEPPELLEKEACIVGELLQOCPRYSWRTSEWKECVSLLEOODPHHWYTG 287
324 ROFPFGSEKECEPEKEPEKCLSGQGVVPCATYGRWTEWTECRVDPLLSQDQKRGNOQT 383
288 PYCGGGIQTREYVYCAQS-----VPAAALRAKEVSRPVEKALCVGPAPLPSOLCNIPCS 343
384 ALCGGGIQTREYVYCVQANENLLSQLSTHKNKESAKPMDLKLCTGPIPNITQLCHIPCPE 443
344 CIVSSKSAWGLCIHENCHEPQCKKGFRTORHVLVESTGPA---CHCPHVESVPCEDPM 400
444 CEVSPMSAWGPTCYENCNDQOGKGFKRRKRIITNEPTGGSGVTCNCPHLLLEAIPCCEPA 503
401 CYRWLASE-GICFPDHGK-CGIGHRIILKAVCONDRGEDVSGSLC-----PVPPPERKSC 453
504 CYDWKAVRLGDCPEPDNGKECGPGTQVQEVVCLNSDGEVDRLCHDAIFPFP-----VAC 558
454 ETPCRMDCVLSEWTEWSSQSCSNKNSDKGQTRSRKTILALAG-BGGKPCPSPQALQHR 512
559 DAPCPKDCVLSTWSTWSSCSHTCSGKTGKQIRARSILAYAGEEGGIRCPNSSALQVR 618
513 LCNDRSCMLHWETSPWGPSCSDTLTALNATIGNWGATCGVGTQTRRVFCVSHVGOV 572
619 SNEHPCTVYHQTFGPGQCIEDTSVSSFNITTTWNGEASCSVGMTKRVLCVRNVGOV 678
573 MTKRCPDSTRPTVPFCPLPCKKDCIVTAFSWBTPCPRMCOAGNATV-KQRYRIIIOEA 631
679 GPKCPESLRPTVPCLLPCKKDCIVTSPDWTSCPSCKEGDSSIRKQSRHVRVILQ 738
632 ANGQSCPTLYEERECEDEVSLCPYRWKPKQWSPCLILVPESVMQGITGSSEACKGLQT 691
739 ANGGRDCTDPLVEEKACEAPQACQSTRWKTHKWRRCQLVPWMSVQDPSGAQEGCGPGRQA 798
692 RAVSCISDDNRSAENMECLKQNGMPLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKS 751
799 RAITCKQDGGQAGIHECLQYAGVPALTQACQIPQDDCQLTSWKFSSCNGDCGAVRT 858
752 RRQLTGKSRKEKQODSLYPLVETELCPDEFTISQYGNWSDCILPEGRREPRHGLRV 811
859 RRTLVGKSKKEKCKNSHLYPLIETQYCPCKDYNAQVPVGNWSDCILPEGKVVELLGMKV 918
812 QADSKECGEGLAFRAVACSDKNRPVDPSPFCSSGVIOBKCVIPFPCKLSDWSWGS 871
919 QGDIKECGGYRYQAWACYDQNGRLVETSRCHSHGYIEACIIPCSCKLSEWNSR 978
872 SSSCGIGVIRSKWLKEKPYNGRCPKLDLKN--QVHEAVPCYSECNQYKSWVEHWSSC 929
979 SKSCGSGVVRSKWLREKPYNGRCPKLDHVNQAQVVEVPVCHSDCNQYLWVTEPWSIC 1038
930 KINNELSLRCGGGTOSRKIRCV-NTADGEGAVDSNLQNDIEIPETQSCSLMCPNECV 988
1039 KVTFTVNRNENCGEVQTRKVRQNTADGPSHEHSDYLCDEEMPLGSRVKLPCPEDCV 1098
989 MSEGWLKWSKQSCDPHTMQTRTHLLR-PSLNSRTCAEDSQVQCLLNENCFQFYNL 1047
1099 ISEWGPWTQVCLPCNQSSPQRASDPIRQPADEGRSCPNAVEPCNLNKNKYHYDYNVT 1158

```

Novel proteins and polynucleotides of secreted proteins useful for treating various diseases e.g. rheumatoid arthritis, cancer, psoriasis, diabetic retinopathy, arteriosclerosis, ischemia or reperfusion injury.

Claim 8; Page 117-123; 145pp; English.

The invention relates to human secreted polypeptides designated LP095, LP191, LP217, LP220, LP221, LP222, LP237 or LP238 and nucleic acid molecules encoding such polypeptides. Novel secreted proteins of the invention are used for treating diseases such as arteriosclerosis, Alzheimer's disease, diabetic retinopathy, psoriasis, pancreatitis, arteriosclerosis, rheumatoid arthritis, colorectal adenomas, severe combined immunodeficiency, ischaemia, carcinomas, haemolytic anaemia, reperfusion injury, neoplasms and cancer especially liver cancer. They are also used for wound healing. Polynucleotides of the invention can be used to generate transgenic animals or knock out animals, which in turn, are useful in the development and screening of therapeutically useful reagents for use in the treatment of diseases associated with LP polypeptide associated activity. They are also used in gene therapy. The present sequence is human LP217 secreted protein. LP217 gene is located on chromosome 7p21-p22

XX Sequence 1624 AA;

Query Match 52.1%; Score 4296; DB 5; Length 1624;  
Best Local Similarity 50.7%; Pred. No. 3e-246;  
Matches 766; Conservative 236; Mismatches 449; Indels 58; Gaps 20;

2 VRTQKLNRTVANEICEHPALPPTFEQACLPICRDCVVFPLPNSKCGKGLQHR 61  
125 IACIQKDKIPAEIDICEYEPKPLEQACLPICQDQIVSEPSAWSECKTGGSLQHR 184  
62 TRAVIAPPLFGGLQCPNLTESACDAPISCPLEEBEYTSFKVPSKCLPHLKEINPS 121  
185 TRHVAVAPPQGGGCPNLTPEFQVCQ---SPCEAEELRSLVHVPWSTCSMPHSRQVROA 241  
122 GRT-----VLDNFNSDNERTFK-HQSVKAAHHSKSWAIEIGYQTRQVSCRT 167  
242 RRRGNKEREKDRSGKYKD--PEARELIIKKRNRRQNRQENKYWDIQGYQTRVWCIN 299  
168 SDGONAMLSLQDSPLPTVQSCIMPDKDSTQSWNSWSPCKTCSGSLLPGRSRRNV 227  
300 KTGKAADLSFCQOEKLPMTQSCVITKEQVSEWSEWSPCKTCHDWSPAGPRVTRTI 359  
228 KMAIGGKECPLELLEKACIVGELLQCCPRYSWRTSEWKEQVSLLEQDPHWHVTG 287  
360 RQFFIGSEKEPEEPKPLQSQDGVVPCATVGRWTEWTECRVPLLSQDXXRGNQT 419  
288 PVOGGGIQTRVYCAQS-----VPAALAAKAEVSRPVEKALCVGPAPLPQLCNIPCS 343  
420 ALCGGGIQTRVYCVQANENLLSLSLTHKNKASKPMDLKLTGPTPNTTQLCHIPCPT 479  
344 CIVSSWSAWGLCIHENCHEPOGKGFTRORHVLMESTGPA---GHCPLHVESVPCEDPM 400  
480 CEVSPWAWGCTYENCNDQGGKGFKLRRRIINEFTGGSGVTGNCPLHLEAIPCEBPA 539  
401 CYRWLASE-GICFPDHGK-CGLGHRILKAYCONDRGEDVSGSLC-----PVPPPPK 453  
540 CYDNKAVRLGDCEPDNGKEGPGTQVQEVVVCINSDEEVDVQRLCRDAI FPIP-----VAC 594  
454 EIPCEMDCVLSEWTEWNSCSCSNKNSDKQTRSRITILALAG-EGKPCPPQALQEH 512  
595 DAPCPKCVLSTWSTWSSCSHTSGKTEGQTRARSILAYAGEEGIRCPNPSALQEV 654  
513 LCNHSCNQLHWETSPWGPCESEDTLVLTALNATIGNGEATCGVIGQTRRVFCVKS 572  
655 SCNEHPCVTHWQCPGQCIEDTSVSSFNITTTWNGEASCSVGMQTRKVICRVNV 714  
573 MTKCPDSTRBETVRPCLPKCKDCIVTASEWTPCPRMQAGNATV-KOSRTRIIQEA 631  
715 GPXKCPESLRBETVRPCLPKCKDCIVTASEWTPCPRMQAGNATV-KOSRTRIIQEA 774  
632 ANGQCECDTLIYERECEDVSLCPVYRWKPKQKSPCIVPESVWQGTGSGEACGKGLQ 691

Db 775 ANGRDCTDPLYEKACEAPQACQSVYKWKHRRCOLVPWSVQDQSPGAQEGCGPQQA 834  
QY 692 RAVSCISDDNRSEMEWECKQTNMGMLLVQECTVPCREDCTFTAMSKFTPCSNCEATKS 751  
Db 835 RAITCRKQDGGQAGIHECLQIAGVPFALTOAQIQDQDQLTSWSKFSKSCNDCGAVRT 894  
QY 752 RRLQTKSRKKEKCDSDILYFVETELCPDCBFIQPYGNWSDCILPEGRREPHRLRV 811  
Db 895 RKRTLVGSKKKECKNSHLYPLIETQYCPDKYNAQPVGNWSDCILPEGKVEVLLGMKV 954  
QY 812 QADSKKEGELRPAVACSDNGRVPDPSCSSGYIOEKCVIPCPDFDCKLSWSSWNSG 871  
Db 955 QGDKEGCGQRYQAMACYQNGRLVETSCNSGHVIEEACIIPCSDCKLSWSSWNSR 1014  
QY 872 SSSCGIGVIRSRKWLKEKPYNGRCPKLDLKN--QVHEAVPCYSECNQYSWVHSSC 929  
Db 1015 SKSCSGGVKVRKWLREKPYNGRCPKLDHVAQAQVYVVPCHSDCNQYLWTEPWSIC 1074  
QY 930 KINNELSLRGGGTQSRKIRCV-NTADGEGGAVDSNLCNQDEIPETQSCSLMCPNECV 988  
Db 1075 KVFVNMRENCGEGVQTRKVRQNTADGSEHVEDYLCDFEEMPLGSRVCKLPCPEDCV 1134  
QY 989 MSEGLMSKCPQSCDPTMORTRHLLR-PSLNSRTCAEDSQVQPCLLNENCFOFQNL 1047  
Db 1135 ISENGPHTQCVLPCNQSFRQSRADPIQPADEGRSCPNAVEKEPCNLNKNKYHYDNT 1194  
QY 1048 EWSQCOLSENAPOCGQVTRLLSCVSDGPKVPMQDQEHNLKPKQMSIPCLVECVVNC 1107  
Db 1195 DWSTCOLSEKAVCGNGIKRMLDCVRSQSGKVDLYKCEALGLEKNWOMNTSCWVECPVC 1254  
QY 1108 QLSGWTAWTECSQTCGHHGRMSRTRFIIMPTQGBGRPCPTLTQECTCPVTYPCYSWVLGN 1167  
Db 1255 QLSWSPWSECSQTCGLTKMIRRTVQFQDGRPCPSLMDQSKPCVPKCYRWQYQ 1314  
QY 1168 WSACKLEGDCGEGVQIRLSJSCMWHSISHAAGRVEDALCGEMPF-----QDSILKOLC 1222  
Db 1315 WSPQVQEAQCGEGTRTNISCVVSDGSADDFSKVVDDEFCADIELIIDGNKNVLESC 1374  
QY 1223 SVPCGDCHLTEWSENSTCELTCDGRSFETVGRSRTFIQSFENQDSCPOVLETR 1282  
Db 1375 SQPCGDCYLDKWSLSLQCTCVNGEDLFGGQVRSRPIQELNQLHCPQMLETK 1434  
QY 1283 PCTGCKYHYTWKASLMNNERTVWCORSQGVNVTGCGSPQARPAARQICIPACRKFPSY 1342  
Db 1435 SCYDQCYEYKWMASAWKSGSRTVWCORSQGVNVTGCGSQAADRSNPPCSPHYSY 1494  
QY 1343 CTQGVCCCEKGYTEIMKNSGFLDYC-----MKVPSGSEDKADVKNLSGKNRPVNSK 1397  
Db 1495 CSETKTCHCEEGYTEVMSNSTLEQCTLIPIVVVLPTWEDKRGDVKT-SRAVHTQPSNP 1553  
QY 1398 IFKG--WSLQPLDPGRVKIWWYGVSGAGFLIMIFLFTSYLVCKXP-KPHQSTPPQOKP 1454  
Db 1554 AGRGRTWELQFPDGRGLKTVYGAAGAVLLIFIVSMIYLACKPKPKPQRQNNRKP 1613  
QY 1455 LTLAYDGDMD 1465  
Db 1614 LTLAYDGDMD 1624

## RESULT 7

AAM39295

ID AAM39295 standard; protein; 1588 AA.

XX

AC AAM39295;

XX

DT 22-OCT-2001 (first entry)

XX

XX Human polypeptide SEQ ID NO 2440.

XX

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
leukaemia.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US034263.

23-DEC-1999; 99US-00471275.

21-JAN-2000; 2000US-00488725.

25-APR-2000; 2000US-00552317.

20-JUN-2000; 2000US-00598042.

19-JUL-2000; 2000US-00620312.

03-AUG-2000; 2000US-00653450.

14-SEP-2000; 2000US-00662191.

19-OCT-2000; 2000US-00693036.

29-NOV-2000; 2000US-00727344.

(HYSB-) HYSBQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

Zhou P, Goodrich R, Dmanac RT;

WPI; 2001-442253/47.

N-PSDB; AA158451.

Novel nucleic acids and polypeptides, useful for treating disorders such

as central nervous system injuries.

Example 4; SEQ ID NO 2440; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the  
encoded polypeptides (AA138442-AA142213) with nootropic,  
immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
in gene therapy. A composition containing a polypeptide or polynucleotide  
of the invention may be used to treat diseases of the peripheral nervous  
system, such as peripheral nervous injuries, peripheral neuropathy and  
localised neuropathies and central nervous system diseases, such as  
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
utilisation of the activities such as: immune system suppression,  
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
assays for receptor activity, arthritis and inflammation, leukaemias and  
C.N.S disorders. Note: The sequence data for this patent did not form  
part of the printed specification

Sequence 1588 AA;

Query Match

Best Local Similarity 52.1%; Score 4292; DB 4; Length 1588;

Matches 765; Conservative 239; Mismatches 449; Indels 58; Gaps 20;

2 VACIOKLNVTVAHEICEFALQPPTEQAELIPCRDCCVVSFPLWNSCKSCGKGLQHR 61

89 IACIQKDKDIPAEIICFYEFPKPLEQAELIPCCQDCIVSFSAWSSECKTCSGGLQHR 148

62 TRAVIAPPLFGGLQCPNLTESRACAPISCPGEBEYTFSLXVGPWSKRLPHLKEINPS 121

149 TRHVAPPFGSGCPNLTEFVQVQ---SSPCEAEELRYSLHVGWSTCSMPHSQVQQA 205

122 GRT-----VLDFNSDSNERVTFK-HQSKAHHSKSWAIGIQTRQVSCTR 167

206 RRGKNKREKDRSGVKDQ--PEARELIKKNRNQRQENKYWDIQIGYQTRVVMCIN 263

168 SDGQNAWLSLCLQDSPLTVQSCIMPKDCQTSQWSWSWSPSCSKTCSGILLPCFRSERNV 227

264 KTGKADLSFCQKELPMTFQSCVITKEQVSEWSEWSPSCSKTCHDMVSPAGTRVTRTI 323

QY 228 XHMAIGGKCEPPELLEKEACIVGELLQOCFPRYSWRTSEWKEQVSLILLQOQPHHVTG 287  
DB 324 RQFFIGSEKCEPFEKEPCLSQDQGVVPCATYWGRTTEWTEBTRDPLLSQDQRNGNQ 383  
QY 288 PVOGGG:QIREVYCAQS---VFAAALRAKESRVPKALCVGPAPLPQNLNIPGSTD 343  
DB 384 ALGGG:QIREVYCVQANELLSQLSTHKNKEASKPMDLKLCTGPIPTNTQLCHIPCPE 443  
QY 344 CIYSSNSAWGLCIHENCHBPQKGGFTRORHVLMESTGPA---GHCPHLVESVPCEPDM 400  
DB 444 CEVSPNSAWGPECTYENCNDQGGKPKLRKRITNEFTGGSGVTGNCPHLEAIPCEPA 503  
QY 401 CYRWLASE-GICFPDGHK-CGLGHRILKAVCONDRGEDVSGSLC---PVPPPPERKSC 453  
DB 504 CYDMKAVRLGDCPEPDNGKECGPTQVQEVVVCINSDEVDRLQCRDAIPFIP-----VAC 558  
QY 454 EIPCRMDCVLSEWTESSQSCSNKNSDKQTRSRITLALAG-BGGKPCPPPSQALQEH 512  
DB 559 DAPCPKDCVLSWTWSSCSHTCSGKTTEGQIRARSILAYAGEEGGIRCPNSSALQEV 618  
QY 513 LCNDSHSCMLHWETSPMGPCSEDTLVTALNATIGWGEATCGVIGIOTRRVFCVKSHVGV 572  
DB 619 SCHEHPCVTVHGTGPGWQCIEDTSVSSFNWTTWGEASCSVGMQTRKVICURVNVGV 678  
QY 573 MTKRCPDSTPRTVPCFLPKCKDCIVTAFSEWTPCPRMQAGNATV-KQSRYRIIIEA 631  
DB 679 GPXKCPESLRPETVRPCLLPCKKDCIVTPYSDWTSFSSCKEGDSIRKQSRHRVILP 738  
QY 632 ANGOECBPDLTYEBERCECVSLCPVVRKPKQKSPCLILVPESVWQITGSSSEACGGLQ 691  
DB 739 ANGRDCTDPLYEKEACEAPQACQSTRWTKHRRRCQLVFWVQQDSPGAQEGCGPRQA 798  
QY 692 RAYSCISDDNRSAMMECLKQTMGPLLVQECVTPCREDCFTFAMSKFTPCSTNCEATKS 751  
DB 799 RAITCRQDQGGQAGIHECLQYAGVPALTAQACIPQDDCQLTSMKSFSSCNGDCGAVR 858  
QY 752 RRQLTKSRKXKCCODSLYPLVETELCPDEFISQPYGNWSDCILPGRRRPRLGLRV 811  
DB 859 RKETLVGSKSKKCKKNSHLYPLIETQYCPCKYNAQPVGNWSDCILPBGKVLGLGMKV 918  
QY 812 QABSKEGEGRLRPAVACSDKNGRPVDPSPCSSGVIQKRCVPCPPDCKLSWSSWMSGC 871  
DB 919 QGDIKEGCGQYRVQMACYDONGELVETSECNHSHVIEACIIPCPSDCKLSWSSWMSGC 978  
QY 872 SSSCGTGVTRIRSKWLKEKPYNGRCPCKLULRN--QVHEAVPCYSCNQYVWVHSSC 929  
DB 979 SKSGSGVKVRSKWLREKPYNGRCPCKLHVNAQAVYEVVPCHSQCNQYLVWTEPWSIC 1038  
QY 930 KINNELSLRCGGGTOSRKIRCV-NTADGEGGAVDSNLGNODEIPPETOSCSLMCPNECV 988  
DB 1039 KTVFVNRNCGEGVQTRKVRQNTADGSEVEDYLCDEMPGLGSRVCKLPCPEDCV 1098  
QY 989 MSBWGLWSKCPQSCDPHTMQRTRHLR-PSLNSRTCAEDSQVQPCLLNENCFQFQYNT 1047  
DB 1099 ISEWGPWTQVLPFCNQSSFRQSRADPIQPADBGRSCPNAVEKPCNLKNCVHYDYNVT 1158  
QY 1048 EWSCTQSENAPCGQVTRITLLSCVSCDKPFYMDQCEQHNLEKPKQMSIPCLVECVNVC 1107  
DB 1159 DWSTQQLSEKAVCGNGIKTRMLDCVRSDDGKVDLYKCEALGLEKNWMTSCMVECPVNC 1218  
QY 1108 QLSGTAWTSCSTCGHGGGMSRTRFIIMPTQGEGRPCPTTELTOEKTCPVTPCYSVLGN 1167  
DB 1219 QLSWSPWSECSQTGTLGKMIRRTVTQPFQDGRPCPSLMDQSKPCPKVPCYRQYQY 1278  
QY 1168 WSACKLEGDCGEGVQIRLSLSCMWHSGSIHAGRVEDALCGEMPF-----QDSILKQLC 1222  
DB 1279 WSPCQVQEAQCCGEGTRNISCVVSDGSADDFSKVDVEEFACDIELIDGNKNWLEESC 1338  
QY 1223 SVPCPGDCHLTENSESTCELTCTIDGRSFETVGRQSRSTRFIIQSPENODSCPOQVLETR 1282  
DB 1339 SQPCPGDCHLTENSESTCELTCTIDGRSFETVGRQSRSTRFIIQSPENODSCPOQVLETR 1398





Y 692 RAVSICSDNRSAAEMWELKQTNMPLLVQECTVPCREDCTFTAWSKTPCSTNCEATKS 751  
b 799 RAITCRKODGGQAGHECLQVAGFPALTAQOIPQDDCOLTSWSKFSNCGDGAVRT 858  
Y 752 RRRQLTGSRRKKEKQDSDPLVLTTELCPDDEFTSQYGNWSDCILPEGRREHRLVR 811  
b 859 RKTLLVGSRRKKEKQDSDPLVLTTELCPDDEFTSQYGNWSDCILPEGRREHRLVR 918  
Y 812 QADSKECEGLFRFRAVACSDKNGRVPDPSFSSSGYIOEKVIPCFFDCKLSDSSWGSC 871  
b 919 QGDIKECGGVYQAMACVDQNGRLVTSRCSNHSYIEACIIPCFSCKLSEWNSRC 978  
Y 872 SSSCGIVRISSKWLKEPYNGRCPCPKLDLX--QVHEAVPCYSECNQYVWVHSSC 929  
b 979 SSSCGGVKVRASKWLREPYNGRCPCPKLDLX--QVHEAVPCYSECNQYVWVHSSC 1038  
Y 930 KINNELSLRSCGGTQSRKIRCV-NTADGEGAVDSNLCNDCEIPEPQSCSLMCPNECV 988  
b 1039 KVTFFVNRNCGGVQTRKVRQCMQNTADGSEHVEDYLCDEPMPGLSRVCKLPCPEDCV 1098  
Y 989 MSEKGLSKPCSDPHWQTRHLR-PSLNSRTCAEDSQVQPCILNENCFQYNLT 1047  
b 1099 ISEWGPWTQCVLPNCQSSFRQADPIRQADDEGRSCNPAVEKEPCNLNKNCHYDYNVT 1158  
Y 1048 EWSCTQSLSENAPCGGVTRILLSCVSDGKPVSDQCEHNLKPEQMSIPCLVECVNVC 1107  
b 1159 DWSTQSLSEKAVCGNGIKTRMLDVRSDGKSDVLYKCEALGKKNWNTSCWECVPNC 1218  
Y 1108 QLSGWNTAECSTQCGHGRMSRTFIIIMPTQGEGRPCPTTELQKTCPTVPCYSWVLGN 1167  
b 1219 QLSDWSPWSECSQTCGLTKKIRRTVTPQFGDGRPCPSLMDQSKPCVPCVRYWYQG 1278  
Y 1168 WSAKLEGGDCGEGVQIESLSCWHSISHAAGRVEDALCGEMPF-----QDSILKQLC 1222  
b 1279 WSPQVQAOCEGRTTRNISCVSVDGADDFSKVDEFCADIELIDGNKNWVLESC 1338  
Y 1223 SVPCGDCHLTEWSEWSRCELTCIDGRSFETVGRSRTFIIISFENQDSQPOQVLETR 1282  
b 1339 SOPCGDCYLKDWSSWSLCLTCVNGEDLFGGQVRSRPIVIOELNQHLCPEQMLETK 1398  
Y 1283 PCTGKCHYTWKASLWNNERTWCQSDGVNVTGCGSPQARPAARQCIIPACKRPFYS 1342  
b 1399 SCYDGCQVEYKMASAWKSGRTVWCQSDGVNVTGCGSPQARPAARQCIIPACKRPFYS 1458  
Y 1343 CTQGVGCEGCEYTIMKSNGLFYDVC-----MKVPGSDKADVKNLGKURPVNSKIH 1397  
b 1459 CSEYKTCHEGTEVMSNSLTLEQTLIPVVUPTMEDKGDVKT-SRAVHPTQPSNP 1517  
Y 1398 IFKG--WSLOPLDPGRVKIYVGVSGAFIMIFLFTSYLVCKKP-KPHQSTPPQOKP 1454  
b 1518 AGRGRTWLPQFPDGRLLKTYGVAGAFVLLIPVSMIYLACKPKPKPORRQNNLKP 1577  
Y 1455 LTLAYDGDLDL 1465  
b 1578 LTLAYDGDADM 1588  
RESULT 9  
BR58459  
D ABR58459 standard; protein; 1588 AA.  
X C ABR58459;  
X T 07-JUL-2003 (first entry)  
X E Human NOV47j.  
X W Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
W immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;  
W antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;  
W diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
W neurodegenerative disorder; Alzheimer's disease; immune disorder;

KW haematopoietic disorder.  
XX Homo sapiens.  
XX WO2003029423-A2.  
PD 10-APR-2003.  
XX 02-OCT-2002; 2002WO-US031358.  
XX 02-OCT-2001; 2001US-0326483P.  
PR 05-OCT-2001; 2001US-0327342P.  
PR 03-OCT-2001; 2001US-0327917P.  
PR 09-OCT-2001; 2001US-0328029P.  
PR 09-OCT-2001; 2001US-0328044P.  
PR 09-OCT-2001; 2001US-0328056P.  
PR 12-OCT-2001; 2001US-0328849P.  
PR 15-OCT-2001; 2001US-0329414P.  
PR 17-OCT-2001; 2001US-0330142P.  
PR 22-OCT-2001; 2001US-0341058P.  
PR 24-OCT-2001; 2001US-0339266P.  
PR 24-OCT-2001; 2001US-0343629P.  
PR 29-OCT-2001; 2001US-0349575P.  
PR 01-NOV-2001; 2001US-0346357P.  
PR 12-APR-2002; 2002US-0371972P.  
PR 12-APR-2002; 2002US-0371980P.  
PR 17-APR-2002; 2002US-0373261P.  
PR 19-APR-2002; 2002US-0373805P.  
PR 23-APR-2002; 2002US-0374738P.  
PR 16-MAY-2002; 2002US-0381101P.  
PR 17-MAY-2002; 2002US-0381635P.  
PR 20-MAY-2002; 2002US-0383830P.  
PR 01-OCT-2002; 2002US-00262839.  
XX (CURA-) CURAGEN CORP.  
PA AIsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
PI Einger SR, Ellemman K, Gerlach VL, Gorman L, Guo X, Ji W;  
PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
PI Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;  
PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
XX WPI; 2003-381625/36.  
DR N-PSDB; ACC72171.  
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
PT dyslipidaemia, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.  
XX Claim 1; Page 264-265; 487pp; English.  
XX The present invention relates to novel human NOV proteins and their  
CC coding sequences (ACC72075-ACC72181 and ABR58459). The NOV  
CC proteins are useful in manufacturing a medicament for treating a syndrome  
CC associated with a human disease. The NOV proteins and coding sequences  
CC may be used to diagnose, treat or prevent metabolic disorders such as  
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
CC disorders such as Alzheimer's disease or Parkinson's disease, immune  
CC disorders, haematopoietic disorders and various dyslipidaemias  
XX Sequence 1588 AA;  
SQ Query Match 52.0%; Score 4284; DB 6; Length 1588;  
Best Local Similarity 50.6%; Pred. No. 1.5e-245;  
Matches 764; Conservative 238; Mismatches 451; Indels 59; Gaps 20;  
QY 2 VRCIQKLNRTVAVNICEHFALQPTBOACLIPOCDVSEFLPWSNCSGCKGLQHR 61  
Db 89 IACIQKDXDIPAEIDICEYFEKPLLEQACLIPOCDQIVSEFSWSECSKTGSGLQHR 148  
QY 62 TRAVIAPPLFGGLQPNLTESRACDAPISCLPGEYEYTFSLKVGWPWSKCRPLHLKEINPS 121

149 TRHVAVPPQGGGCPNLTEFQVCQ---SSPCAEELRYSLHVGPMSTCSMPHSRQVRQA 205  
122 GRT-----VLDFNSDSNERVTFK-HOSYKAHHSHKSAIEIGYOTROVSCTR 167  
206 RRGKKNKEREKDSKVGD---PEARLIIKKKNRPNQENKYWDIQIGYOTREVMCIN 263  
168 SDQONAMLSLCLQDSPLTVQSCIMPKDCEBTSOWSPSCSKTCSGSLLPFRSRSRV 227  
264 KTGKAADLSFCQCKELPMTFQSCVITKECQVSEWSEWSPCKTCHDMVSPAGTRVRTTI 323  
228 KMAIAGGKECPLELLKEACIVGEILLQOCPRYSWRTSEWKECQVSLLEQQDPHHVYG 287  
324 RQPISEKECPFEKEKPLCSGDGVPCATYGNRTTEWTECRVDPLLSQQDKRGNOT 383  
288 FVCGGIQTRETVCAQS----VFAAALRAKEVRSRVEKALCVGAPLPSQLCNT:PCSTD 343  
384 ALCGGIQTRETVCVQANENLLSLXTHKNKEASKPMDLKLCTGPINPTTXLCH:PCPTE 443  
344 CIYSSSAMGLCIHENCHEPQGGKGRTRQRHVLMESTGPA---GHCPLVRSVCEBPM 400  
444 CEVSPSAMGCPCTYENCNDXXKKGKFKLRKRRTITNEPTGSGVGTGNCPLHLEAICEEPA 503  
401 CYRWLASE-GICFPDHRGK-CGLGHRILKAVCQNDRGEDVSGSLC-----PVPPPPPERKSC 453  
504 CYDWKAVRLGDCPEPDNGKEGPGTQVQEVVVCINSDEEVDRQLCRDAIFPIP-----VAC 558  
454 EICRMDCVLSEWTEWSSSCSOSCNKNSDKQTRSTIILALAG-EGGKPCPPSQALQHR 512  
559 DAFCPKDCVLSTWSTWSSSHTCSGKTGKQIRARSILAYAGEEGGIRCPNSSALQEV 618  
513 LCNHDSQWQJHWETSPWGPCSEDTLTALNATIGMNGEATCGVIGITRRVFCVKSHVGOV 572  
619 SCNEHPCVTVHWTGPNQGCIEDTSVSSNTTWNAGEASCVSQWTRKVICVRVNVGOV 678  
573 MTRCPDSTRPTEVRPCFFLPCKKDCIIVTAFSEWTPCPRMCOQAGNAV-KQSYRIRIIQEA 631  
679 GPKKCPESLREPVRPCLLPCKKDCIIVTSPDWTSCPPSSCKEDGSIRKQSHRVIQLP 738  
632 ANGGQECPTLYEERCEDEDVSLCPVVRWPKQWSPCILVPESWQGITGSSAEACGKQJT 691  
739 ANGRCTDPLYEKEACEAPQACQSRWTKHRRRCOLVPSVQQSPQAGCGGPGQA 798  
692 RAVSCISDDNRSAMMECLKQTMGPLLVQECTVPCREDCTTAMSKFTPCSTNCEATKS 751  
799 RAITCRKQDGGQAGIHECLOYAGVPALVQACQIPQDDCQLTSKSKFSCNGDCGAVRT 858  
752 RRBLTKSKKKEKQDSLDLYLVEITELCEDEPIISQPYGNWSDCILPGRPREHRLAV 811  
859 RKRTLVGSKSKKKEKKNSHLYPIETQYCPCKYNAQPVGNWSDCILPGRKVEVLLGMKV 918  
812 QADSKEGGLREFRAVACSDKNGRPVDPSPFCSSGYIQEKVIPCFFDCKLSDWSW3SC 871  
919 QGDIKECGQYRYQAMACYDQNGRLVETSRCSNHSYIEACIIPCPSDCKLSEWNSW3SC 978  
872 SSSCGIGVIRSRKWLKXKPNGBRCPKLDLKN--QVHEAPVCYECNOYSWVVEHW3SC 929  
979 SKSCGSGVKVRSKWLREKPNYNGRCPKLDHVNQAQVYEVFCHSDCNOYLWVTFW3SC 1038  
930 KINNELRLSCGGGTQSRKIRCV-NTADGEGGAVDSNLNODBIPEBETCSILMCPNECV 988  
1039 KVTFVNNRENCGEVQTRKVRQNTADGEPSEVEDYLCDPEMPLGSRVCKLPCEDCV 1098  
989 MSWGLWNSKCPQSDPHMTQRRTHLLR-PSLNSRTCAEDSQVQCLLNENCFQOYNLT 1047  
1099 ISEMGPTQVCLPNCSSFPQRSADPIQPADEGRSCPNAVEKPCNLNKNCHYDYNVT 1158  
1048 EWSTCOLSENAPCGGVTRTELLSCVSDGKPVSMDOCEQHNLEKPBORMSIPCLVRVNC 1107  
1159 DWSTCQSEKAVONGIKIKNLDCVSDGSKVDLKYCBALGLEKWKQMTSCWKEPCVNC 1218  
1108 QLSGWTATWECSTQCGHGRMSRTRFIIMPTQEGRPPCTELTQECTPVTFCYSWLVGN 1167  
1219 QLSWSPSWSECSQTCGLTGKWRRTVTQPFQDGRFCPSLMDQSKFCPVKFCYRWQYQ 1278

1168 WSACKLGGGCGEGVQIORSLSQMVHSGSISHAAGRYVEDALCGEMP-----QDSILKQLC 1222  
1279 WSPQVQEAQCGEGTRTNISCVVSGSADDFKVDDEFCADIELIIDGNKMWLEESC 1338  
1223 SVPCGCHLTXEWSMTCELTCTIDGRSFETVQRQSRSTFIIOSEPNQDSCPQVLETR 1282  
1339 SQPCGDCYLKDWSSWSLCLQITCVNGEDLGFGGIGVRSRPVITQELENOHLCEQMLETK 1398  
1283 PCTGGKCVHYTWKASLWNNNERTWCORSQVNVVTGSCSPQARPAAIROCIACRKPFSY 1342  
1399 SCYDGCYEFKWSAWKSGSRVWCORSQVNVVTGCLVMSQPDADRSNPPCSPSHSY 1458  
1343 CTQGVGCGCEKGYTEIMKNSGFLDYC-----MKVPGSEDKKADVKNLSGKRRVNSKIHD 1397  
1459 CSETKTCHCEGYTEVMSNSNSTLEQCTLIPIVVLPTMEDKRGDKT-SRAVHPTQPSNP 1517  
1398 IFKG--NSLOPLDPDGEVKIWWVSGSAGELIMFLIFTSYLKCKP-KPHOSTPPOQKP 1454  
1518 AGRGRTWLPQFPFGDGLKKTWYVGAAGAFVLLIFIVSMIYLACKKPKPKPQRQNNRLKP 1577  
1455 LTLAYDGLDM 1465  
1578 LTLAYDGDAD 1588

## RESULT 10

ABR58453  
ID ABR58453 standard; protein; 1545 AA.

AC ABR58453;

XX XX 07-JUL-2003 (first entry)

XX Human NOV47d.

XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;  
antiparkinsonian; antilipemic; gene therapy; metabolic disorder;  
diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
neurodegenerative disorder; Alzheimer's disease; immune disorder;  
haematopoietic disorder.

XX Homo sapiens.

XX WO2003029423-A2.

XX 10-APR-2003.

XX 02-OCT-2002; 2002WO-US031358.

XX 02-OCT-2001; 2001US-0326483P.

XX 05-OCT-2001; 2001US-0327342P.

XX 09-OCT-2001; 2001US-0327917P.

XX 09-OCT-2001; 2001US-0328029P.

XX 09-OCT-2001; 2001US-0328044P.

XX 09-OCT-2001; 2001US-0328056P.

XX 12-OCT-2001; 2001US-0328849P.

XX 15-OCT-2001; 2001US-0329414P.

XX 17-OCT-2001; 2001US-0330142P.

XX 22-OCT-2001; 2001US-0341058P.

XX 24-OCT-2001; 2001US-0339266P.

XX 24-OCT-2001; 2001US-0343629P.

XX 29-OCT-2001; 2001US-0349575P.

XX 01-NOV-2001; 2001US-0346357P.

XX 12-APR-2002; 2002US-0371972P.

XX 12-APR-2002; 2002US-0371980P.

XX 17-APR-2002; 2002US-0372612P.

XX 19-APR-2002; 2002US-0373805P.

XX 23-APR-2002; 2002US-0381101P.

XX 17-MAY-2002; 2002US-0381635P.

XX 29-MAY-2002; 2002US-0383830P.



KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytotatic; nootropic; neuroprotective; dyslipidaemia;  
KW antiparkinsonian; antilipidemic; gene therapy; metabolic disorder;  
KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
KW neurodegenerative disorder; Alzheimer's disease; immune disorder;  
KW haematopoietic disorder.  
XX  
OS Homo sapiens.  
XX WO2003029423-A2.  
XX  
XX 10-APR-2003.  
XX  
XX 02-OCT-2002; 2002WO-US031358.  
XX  
XX 02-OCT-2001; 2001US-0326483P.  
XX 05-OCT-2001; 2001US-0327342P.  
XX 09-OCT-2001; 2001US-0327917P.  
XX 09-OCT-2001; 2001US-0328029P.  
XX 09-OCT-2001; 2001US-0328044P.  
XX 09-OCT-2001; 2001US-0328058P.  
XX 12-OCT-2001; 2001US-0328849P.  
XX 15-OCT-2001; 2001US-0329414P.  
XX 17-OCT-2001; 2001US-0330142P.  
XX 22-OCT-2001; 2001US-0341058P.  
XX 24-OCT-2001; 2001US-0339266P.  
XX 24-OCT-2001; 2001US-0343629P.  
XX 29-OCT-2001; 2001US-0343575P.  
XX 01-NOV-2001; 2001US-0346357P.  
XX 12-APR-2002; 2002US-0371972P.  
XX 12-APR-2002; 2002US-0371980P.  
XX 17-APR-2002; 2002US-0373261P.  
XX 19-APR-2002; 2002US-0373805P.  
XX 23-APR-2002; 2002US-0374738P.  
XX 16-MAY-2002; 2002US-0381101P.  
XX 17-MAY-2002; 2002US-0381635P.  
XX 29-MAY-2002; 2002US-0383830P.  
XX 01-OCT-2002; 2002US-00262839.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Alsebrook JP, Anderson DW, Boldog PL, Burgess CE, Carterton E;  
XX Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
XX Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
XX Rothenberg ME, Shinkets RA, Smithson G, Spytek KA, Taupier RJ;  
XX Vernet CAM, Voss EZ, Zerhusen BP, Zhong M;  
XX  
XX WPI: 2003-381625/36.  
XX N-PSDB; ACC72167.  
XX  
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
XX treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
XX dyslipidemia, and in chromosome mapping, tissue typing or  
XX pharmacogenomics.  
XX  
XX Claim 1; Page 260; 487pp; English.  
XX  
XX The present invention relates to novel human NOV proteins and their  
XX coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
XX proteins are useful in manufacturing a medicament for treating a syndrome  
XX associated with a human disease. The NOV proteins and coding sequences  
XX may be used to diagnose, treat or prevent metabolic disorders such as  
XX diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
XX disorders such as Alzheimer's disease or Parkinson's disease, immune  
XX disorders, haematopoietic disorders and various dyslipidaemias  
XX  
XX Sequence 1549 AA;  
XX  
XX Query Match 49.7%; Score 4095.5; DB 6; Length 1549;  
XX Best Local Similarity 50.2%; Pred. No. 2.2e-234;  
XX Matches 730; Conservative 230; Mismatches 437; Indels 57; Gaps 19;  
XX  
XX 2 VRCIQKLNRTVVANIECHFAFPTEQAQLPCPRDCVSEFLFWSNCSKCGKQLQHR 61

Db 91 IACIQKDKDIPAEIDIIIEYFEKPLEAOACLPQQDCIVSEFSWSECSKTCGSGLOHR 150  
Qy 62 TRAVIAPPLFGGLQCPNLTESACDAPISCPUGEEYVTSKLVGWPWSKRLPHLEINPS 121  
Db 151 TRHVAPPOFGSGGCPNLTFFQVCO---SSPCEABELYSLVHVGWSTCSMPHSQVRQA 207  
Qy 122 GRT-----VLDNDSNERNVTFK-HQSYKAHHHSKSWAIEIGYQTRQVSCTR 167  
Db 208 RRRGKNKEREKDRSGVKD--PEARELIKKNRNRQRNQENKYWDIQIGYQTRVVMCIN 265  
Qy 168 SDGONAMLSLCLQDSFPLTVQSCIMPKDCETQSMWSWSPSCSKTCHSGSLLPGFRSRNV 227  
Db 266 KTGKAADLSFCQOEKLPMTFQSCVITKEQVSENERSPCSKTCHEMVSAGTRVTRI 325  
Qy 228 KMAIGGKCEPPELLEKEACIVEGELLQOCPRYSWRTSEWKEQVSLLEQQDPHHVYG 287  
Db 326 RQPPIGSEKEPEFEKEPCLSQGGVWPVCATYGWRTTEWTECRVDPILLSQQKRGNT 385  
Qy 288 PVGCGGIOTREYVCAQS-----VPAALRAKEVSPPEKALCVGPAPLPSCLCNTPCSTD 343  
Db 386 ALCGGGIOTREYVCOANENLLSQJSTHKNKEASKPMDLKLCCTGPIPNMTQJCHIPCPT 445  
Qy 344 CIVSSWSAMGLCIHENCHEPOGKGFRTQRHVLMESTGPA---GHCPLHVESVPCEDPM 400  
Db 446 CEVSPWSAMGPTCYENCNDQCKGFKLRKRRIITNEPTGGSGVTGNCPHLLAIPCEBPA 505  
Qy 401 CYRWLASE-GICFPDHGK-CGLGHRILKAVCONDRGEDVSGSLC-----PVPPPERKSC 453  
Db 506 CYDMKAVRLGNCEPDNGKECGFGTVQVEVVCINSDEGEVDQLCRDAIFPIP-----VAC 560  
Qy 454 ETPCRMDCVLSWETWSSCSQCSNKNSDGKTRTITIALAG-EGGKPCPPSPQALQHR 512  
Db 561 DAPCPKCVLSTWSTWSSCSHTCSGKTTEGKQIRARSILAYAGEGGIRCPNSSALQEV 620  
Qy 513 LNDHSCMLHMETSPWGPCSEDTLVLTALNATIGNGATCGVGIQTRVFCVKSHVGQV 572  
Db 621 SCNEHPTVYHMQTGPWGQCIEDTSVSESNITTTWNGEASCSVGMQTRKVICRVNVGQV 680  
Qy 573 MTKRCPDSTRPETVRPCPLPCKKDCIVTAFSEWTPCPRMCOAGNATV-KQSYRIIIIEA 631  
Db 681 GPKCPESLRPETVRPCLLPCKKECIVTYSWTSWTSWTSWTSWTSWTSWTSWTSWTSWTS 740  
Qy 632 ANGQCEPDTLYEERECEPDVSLCPVYRWKPKQWSPCIIIPVPSWQGITGSSEACGKQ 691  
Db 741 ANGGRDCTDPLYEERACEAPQACQSYRWKTHKRRRCQLVPWSVQDSPGAQEGCGPRQA 800  
Qy 692 RAVSCISDDNRSAMMECLKQINGMELLVQECTVPCREDCTFTAMSKETPCSTNCEATKS 751  
Db 801 RAITCKKQGGAGHIECLQYAGVPALTAQCIPODDCQUTSWKFSKSCNGDCGAVRT 860  
Qy 752 RRRQLTGSRKKEKQDSLYPLVETELCPDCEFTISQYGNWSDCILPEGREPHRLVR 811  
Db 861 RKETLVGSKSKKECKNSHLYPLIETQVCPDKYNAQPVGNWSDCILPEGVVLLGMKV 920  
Qy 812 QADSKECEGLPRAVACSDKNRPVDPSPFCSSSYIQEKVIPCPFPCKLSDNWSWSC 871  
Db 921 QGDIKECGGYRYQAMACYDQNGRLVETSRCSHGYIEACIIPCPCDKLSEWNWSRC 980  
Qy 872 SSSCGIGVIRSKWLKEKPYNGRCPKLDLKN--QVHEAVPCYSECNQYSWVVEHWSSC 929  
Db 981 SKSCGSGVAVRSKWLREKPYNGRCPKLDHVNQAVYEVVPCVHSDCNQYLWTFPWSIC 1040  
Qy 930 KINNELRSURCGGTQSRKIRCV-NTADGEGAVDSNLQNDIEIPPTQSCSLMCPNCEV 988  
Db 1041 KVTVMNRENCGEVQTRKVRQWNTADGSEHVEDYLCDEPEMPLGSRVCKLPCPEDCV 1100  
Qy 989 MSEWGLWSKSCPOSQDPHTMQRTRHLR-PSLNSRTCAEDSOVQCLLNENCFQYNLT 1047  
Db 1101 ISEGWPTQCVIPCNQSSPRQSRADPIRQPADEGRSCFNAVEKEFCNLKNKYHYDNYT 1160  
Qy 1048 EWSVTOOLSENAPCGGVTRTLSSVCVSDGKPVSMQDQCEHNLKLEKQFMSIPCLVECVNVC 1107

1161 DWSTCOLSEKAVCGNGIKTRMLDCVRSDCKSVLDKYCEALGLEKNQWNTSCWBCPVNC 1220  
1108 QLSGWTATWECSTCHGGRMSRTRFIMPTQGEGRPCPELTQKTCVTPCYSWVLGN 1167  
1221 QLSWSPWSECSCTCLGKMIARRTVTPFQDGRGPCPSLMDQSKPCFVPCYRWQYG 1280  
1168 WSACKLEGGDCGEGVQIRSLSCWVHSGSISHAAGRVEDALCGEMPF- ----QDSILKQLC 1222  
1281 WSPCQVQEAQCGEGTRNISCVSVDGSDGADFSKVDBEFCADIELIIDGNKNWVLESC 1340  
1223 SVPCGDCHLTWSEWSTCELICIDGRSPETVGRSRRTFIIQSFENQDSCPOQVLETR 1282  
1341 SPCPCGDCVLDKWSWSLCLTQVANGEDLFGGIGVRSRPIVIOELNQHLCPEQMLETK 1400  
1283 PCTGGKCYHYTWKASLWNNERTVWCQRSDGVNVTGGCSQARPAARQCIQACRKPFSY 1342  
1401 SCVDGQCYEYKMAWASAKGSSRTVWCQRSDGINVTGGCLWMSQPADRSCHNPPCSQPHSY 1460  
1343 CTQGGVCGCEKGYTEIMKNSGFLDYC- ----MKVPGSDKXADVKNLSGKORPVSNIHD 1397  
1461 CSETKCHCEGYTEYMSNSTLEOCTLIPVVVLPVTMBEXRGDVKT-SRAVHPTOPSSNP 1519  
1398 IFKG--WSLOPLDP 1409  
1520 AGRGRTWFLQPGP 1533  
RESULT 12  
ABR58454  
[D ABR58454 standard; protein; 1490 AA.  
[C ABR58454;  
[X ABR58454;  
[T 07-JUL-2003 (first entry)  
[E Human NOV47e.  
[W Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
[W immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;  
[W antiparkinsonian; antipapemic; gene therapy; metabolic disorder;  
[W diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
[W neurodegenerative disorder; Alzheimer's disease; immune disorder;  
[W haematopoietic disorder.  
[S Homo sapiens.  
[X WO2003029423-A2.  
[D 10-APR-2003.  
[F 02-APR-2002; 2002WO-US031358.  
[X 02-OCT-2001; 2001US-0326483P.  
[R 05-OCT-2001; 2001US-0327342P.  
[R 09-OCT-2001; 2001US-0327917P.  
[R 09-OCT-2001; 2001US-0328029P.  
[R 09-OCT-2001; 2001US-0328044P.  
[R 09-OCT-2001; 2001US-0328056P.  
[R 12-OCT-2001; 2001US-0328849P.  
[R 15-OCT-2001; 2001US-0329414P.  
[R 17-OCT-2001; 2001US-0330142P.  
[R 22-OCT-2001; 2001US-0341058P.  
[R 24-OCT-2001; 2001US-0339266P.  
[R 24-OCT-2001; 2001US-0343629P.  
[R 29-OCT-2001; 2001US-0349575P.  
[R 01-NOV-2001; 2001US-0345357P.  
[R 12-APR-2002; 2002US-0371972P.  
[R 12-APR-2002; 2002US-0371980P.  
[R 17-APR-2002; 2002US-0373211P.  
[R 19-APR-2002; 2002US-0373805P.  
[R 23-APR-2002; 2002US-0374738P.  
[R 16-MAY-2002; 2002US-0381101P.  
[R 17-MAY-2002; 2002US-0381635P.

29-MAY-2002; 2002US-0383830P.  
01-OCT-2002; 2002US-00262839.  
(CURA-) CURAGEN CORP.  
Alsbrook JP, Anderson DW, Boidog FL, Burgess CE, Catterton E;  
Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
Rothenberg WE, Shinkets RA, Smithson G, Spytek KA, Taupier RO;  
Vernet CAM, Voos EZ, Zerhusen BD, Zhong M;  
WPI; 2003-381625/36.  
DR N-PSDE; ACC72166.  
XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
dyslipidaemia, and in chromosome mapping, tissue typing or  
pharmacogenomics.  
XX Claim 1; Page 258-259; 487pp; English.  
XX The present invention relates to novel human NOV proteins and their  
coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
proteins are useful in manufacturing a medicament for treating a syndrome  
associated with a human disease. The NOV proteins and coding sequences  
may be used to diagnose, treat or prevent metabolic disorders such as  
diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
disorders such as Alzheimer's disease or Parkinson's disease, immune  
disorders, haematopoietic disorders and various dyslipidaemias  
XX Sequence 1490 AA;  
SQ Query Match 48.9%; Score 4033; DB 6; Length 1490;  
Best Local Similarity 48.1%; Pred. No. 1.1e-230;  
Matches 724; Conservative 224; Mismatches 414; Indels 142; Gaps 18;  
QY 2 VRGLOKLRVTVANEICEHFALQPPTEOACLPDPCDVSEFLPWSNCSKGGCKLOHR 61  
DB 89 IACIQKDKOIPAEDIIICEYFEPCLLEQACLPQOQCIIVSEFSAWSECKTCGSLQHR 148  
QY 62 TRAVIAPPLFGGLQCPNLTESACDAPISCPICEEYTFSLKVGPMKCRLLPLKLNPS 121  
DB 149 TRHWVAPPQGGSGCPNLTEFQVCQ--SSPCEABELRYSLHVGPMSTCSMPSHRQVA 205  
QY 122 GRT-----VLDPSDSNERTVTK-HQSYKAHHSHKSWAIEIGYQTRQVSCTR 167  
DB 206 RRRGKXKEREKDRSKGVKD--PEARELIKKNRNRQENKYWDIQGYQTRVWCIN 263  
QY 168 SDGONAMLSLCLQDSPLTVQSCIMPKDCETSSWSPSCSKTCRSGSLLPGRSSRN 227  
DB 264 KTGKAADLSFCQEQKLPMTFQSCVITKEQVSESWSPSKTCHDMVSPAGTRVTRI 323  
QY 228 KHAIGGGKECPLEKBAIVEGELLQOCPRYSWTSEWKECQVSLLEQDPHWHVTG 287  
DB 324 RQPPIGSEKECEPEFEKEPCLSQDQGWFCATYGTWTEWTECRVDPPLLSQDQKRGNT 383  
QY 288 FVCGGGIQTREVVCAOS-----VPAALRAKEVSRVEKALCVGPAPLSQLCNIPCS 343  
DB 384 ALCGGGIQTREVVCAQANENLLSQLSTHKNKASKPMDLKLTGPIPTQLCHIPCPE 443  
QY 344 CIVSSWSAWGLCIHENCHEPQGGKGRTRQRHYLMESTGPA---GHCPLHVESVPCED 400  
DB 444 CEVSPWSAWGFCYENCNDQOQKKGKFLKRRTITNEPTGGSGVGTGNCPLLEAIPCE 503  
QY 401 CYRWLASE-GICFPDHGK-CGLGHRILKAVQNDRGEDVSGSLC-----PVPPPRK 453  
DB 504 CYDWKAVRLGDCPEFNGKECGPGTQVEVVCINSDEGEVDRLCRDAIFPIP----- 558  
QY 454 EIPCRMDCVLSEWTEWSSQSCSNKNSDKOTRSRTILALAG-EGGKCPSPSOALQHR 512  
DB 559 DAPCPKDCVLSWTWSSSHCSHCSGKTGKIRASILAYAGEEGGICPNSSALXEV 618  
QY 513 LCNDHSCMLHWTSPWGPCSEDTLVTALNATIGWNEATCGVGIOTRRVFCVKGHVQV 572

```
db 619 SCNEHCTVYHWXTPGACXIEDTSVSSNTTWTNGEASCSVGMXTRKVICVRNVGV 678
2y 573 MTKRCDSTPRTVTRPCFLPKCKDCIIVTAFSEWTPCFRMCQAGNAV-KOSRYRIIIQEA 631
db 679 GPKKCPESLRFETVRPCLLPKCKDCIIVTAFSEWTPCFRMCQAGNAV-KOSRYRIIIQEA 738
2y 632 ANGQECDFLYEERCEDEVSLCPVVRWPKQKWSPCILVPESVMOGITSSACGKGLQT 691
db 739 ANGRCCTDLYEKEACEAPQACQSRWTKHKKRQLVPWSVQDSPAQCGGPGRA 798
2y 692 RAVSCISDDNRSABMECLQKQINGMELLVQECTVPCREDCTFTANSKFTPCSTNCEATKS 751
db 799 RAITCRKQDGGQAGIEHCLQYAGVPAITACQIPQDDCQLTWSKFSKSCNGDCGAVRT 858
2y 752 RRRLTGSKSEKKEKCDSDLYPLVETELCPDEFISQPYGNWSDCILPEGRREPHGLRV 811
db 859 RKTTLVGSKKKCKKNSHLYPIETQYCPDKYNAQPVGNWSDCILPEGRREPHGLRV 918
2y 812 QADSKEGGLRFRAVACSDKNGRPVDPSPSCSSGYIOEKCVIPCFDCKLSDWSW3SC 871
db 919 QGDIKECGQYRYQMACYDQNGRLVETSRNCNLSHGVIEEACIIPCFSDCKLSEWSN3SC 978
2y 872 SSSCGIGVTRSKWLKEKPNRGPCKLDLKN--QVHEAVPCYSECNQYSWVEHWSSC 929
db 979 SKSCGSGVKYRSKWLREKPNRGPCKLDLKN--QVHEAVPCYSECNQYSWVEHWSSC 1038
2y 930 KINNELSLRCGGTQSRKIRCV-NTADGEGGAVDNLNQDRIPEPTQSCSLMCFNECV 988
db 1039 KVTFVNNRENGEGVQTRKVRCKMONTADGSEHVEDYLCDEMPVLSRVCKLPCFEDCV 1098
2y 989 MSEWGLWSKCPQSCDPTMORRTHLLR-PSLNSRTCAEDSQVQPCLLNENCFOYNT 1047
db 1099 ISEKPTQVCLPCNQSSFPQADPIRQADRESCPNNAVEKPCNLKNCYHYDNYT 1158
2y 1048 EWSSTCOLSENAPCGQVTRLLSCVSDGKPVSMDOCEQHNLEKXPQMSIPCLV 1107
db 1159 DASTCOLSEKXAVCGNGIKTRMLDCVRSKSDGKSVDLKYEALGLEKQWMTSCNVECPVNC 1218
2y 1108 QLSGWTAWTSCQTCGHRMSRTRFTIMPTQGEGRPCPTLTQETKTCPTVPCYSWVLGN 1167
db 1219 QLSDWSPWSECSQTCGLTKMIRRTVTPQFGDGRPCPSLMDQSKPCPKYRQWYQ 1278
2y 1168 WSACKLGGGCGEGVQIRSLSCMVHSGSI SHAAGRVEDALCGEMPF-----QDSILKOLC 1222
db 1279 WSPQVGEAQCGEGRTRNISCVVSDGSADDFKVDDEFCADIELIIDGNKNWLESC 1338
2y 1223 SVPCPGCHLTENSEWSTCELTICIDGRSFETVGRQSRRTFIIQSPENQDSCPCQVLETR 1282
db 1339 SQPCPGCYLKDWSWSLCLQTCVNGEDLFGGIGVRSRPVILQELNHLCPQMLETK 1398
2y 1283 PCTGGKCVHYTWKASLWNNERTVWCQSDGVNVTGCSQPARPAATROCI PACRKPF 1342
db 1399 SCYQGCVEYKMASAWKSGSRVWCQSDGVNVTGCSQPARPAATROCI PACRKPF 1433
2y 1343 CTQGGVCGCEKGYTEIMKSNGLFDYCMKVPGEDEKADVNKLSGKNRPVNSKIHDIFKG 1402
db 1434 ----- 1433
2y 1403 SLQPLDPDGRVKIWIYGVSGAFILMIFLFTSVLVCKPK-KPHQSTPPQCKPLTLAYDG 1461
db 1434 -----DGRUKTWYGYAAGAFVLLIFTVSMYIACKPKPKQRRQNNRLKPLTLAYDG 1486
2y 1462 DLDM 1465
db 1487 DADM 1490
```

RESULT 13  
AAO15279  
ID AAO15279 standard; protein; 484 AA.  
XX  
AC AAO15279;

```
XX 16-SEP-2002 (first entry)
DT Human large protein 53-24.
XX Human; large protein 53.24; embryonic development deformity;
XX protein metabolism disturbance; tumour; immunologic system disturbance.
XX Homo sapiens.
XX CN1333246-A.
XX 30-JAN-2002.
XX 07-JUL-2000; 2000CN-00117029.
XX 07-JUL-2000; 2000CN-00117029.
XX (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.
XX Mac Y, Xie Y;
XX WPI; 2002-305572/35.
XX N-PSDB; AAL43703.
XX New human large protein 53.24 and encoding polynucleotide, useful for
XX treating embryonic development deformity, protein metabolism disturbance,
XX tumor and immunologic system disturbance disease.
XX Claim 1; Page 29-30 (Disclosure); 36pp; Chinese.
XX The invention comprises the amino acid and coding sequence of the human
XX large protein 53.24. The 53.24 DNA and protein sequences of the invention
XX are useful for treating embryonic development deformity, protein
XX metabolism disturbance, tumour and immunologic system disturbance. The
XX present amino acid sequence represents the human large protein 53.24
XX
XX Sequence 484 AA;
Query Match 33.1%; Score 2729; DB 5; Length 484;
Best Local Similarity 99.8%; Pred. No. 7.7e-154;
Matches 483; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 982 MCPNECVNSEGLWSKCPQSCDPTMORRTHLLRPSLNSRTCAEDSQVQPCLLNENCFO 1041
DB 1 MCPNECVNSEGLWSKCPQSCDPTMORRTHLLRPSLNSRTCAEDSQVQPCLLNENCFO 60
QY 1042 FOYNLTENSTCOLSENAPCGQGVTRLLSCVSDGKPVSMDOCEQHNLEKXPQMSIPCLV 1101
DB 61 FOYNLTENSTCOLSENAPCGQGVTRLLSCVSDGKPVSMDOCEQHNLEKXPQMSIPCLV 120
QY 1102 ECVNVCQLSGWTAWTSCQTCGHRMSRTRFTIMPTQGEGRPCPTLTQETKTCPTVPCY 1161
DB 121 ECVNVCQLSGWTAWTSCQTCGHRMSRTRFTIMPTQGEGRPCPTLTQETKTCPTVPCY 180
QY 1162 SWLGNWSACKLGGDCGEGVQIRSLSCMVHSGSI SHAAGRVEDALCGEMPFQDSILKQL 1221
DB 181 SWLGNWSACKLGGDCGEGVQIRSLSCMVHSGSI SHAAGRVEDALCGEMPFQDSILKQL 240
QY 1222 CSVPCPGCHLTENSEWSTCELTICIDGRSFETVGRQSRRTFIIQSPENQDSCPCQVLETR 1281
DB 241 CSVPCPGCHLTENSEWSTCELTICIDGRSFETVGRQSRRTFIIQSPENQDSCPCQVLETR 300
QY 1282 RPCTGGKCVHYTWKASLWNNERTVWCQSDGVNVTGCSQPARPAATROCI PACRKPF 1341
DB 301 RPCTGGKCVHYTWKASLWNNERTVWCQSDGVNVTGCSQPARPAATROCI PACRKPF 360
QY 1342 YCTQGGVCGCEKGYTEIMKSNGLFDYCMKVPGEDEKADVNKLSGKNRPVNSKIHDIFKG 1401
DB 361 YCTQGGVCGCEKGYTEIMKSNGLFDYCMKVPGEDEKADVNKLSGKNRPVNSKIHDIFKG 420
QY 1402 WSLQPLDPDGRVKIWIYGVSGAFILMIFLFTSVLVCKPKPKQRRQNNRLKPLTLAYDG 1461
```



421 WSLQPLDPGRVKIWKYVGVGGAFILMIFLFTSYLVCKKPKFHQSTPPQKFLTLAYDG 480  
1462 DLDW 1465  
481 DLDW 484  
RESULT 14  
ADB64774  
ID ADB64774 standard; protein; 933 AA.  
AC ADB64774;  
XT 04-DEC-2003 (first entry)  
DE Human protein encoded by clone NT2RP70056690.  
CW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
CW cell regeneration; membrane protein; signal transduction-related protein;  
CW transcription-related protein; osteoporosis; neurological disease;  
CW cancer; tumour.  
CS Homo sapiens.  
XS  
SN EP1308459-A2.  
PD 07-MAY-2003.  
XX 28-MAR-2002; 2002EP-00007401.  
XX 05-NOV-2001; 2001JP-00379298.  
XX 25-JAN-2002; 2002US-00350978.  
PA (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIO TECHNOLOGY.  
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
WPI; 2003-450961/43.  
N-PSDB; ADB62804.  
New polynucleotides and polypeptides, useful for developing a diagnostic  
marker or medicines for regulation of their expression and activity, or  
as targets of gene therapy.  
Claim 1; Page; 222pp; English.  
The invention discloses a polynucleotide comprising a sequence selected  
from 1970 fully defined nucleotide sequences which encode novel  
polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
or its partial peptide, an antibody binding to the polypeptide or peptide  
of the polynucleotide, immunologically assaying the polypeptide or  
peptide of the polynucleotide by contacting the polypeptide or peptide  
with the antibody of the encoded protein, and observing the binding  
between the two, a transformant carrying the polynucleotide in an  
expressible manner and an antisense polynucleotide. The oligonucleotide  
is useful as a primer for synthesising the polynucleotide, or as a probe  
for detecting the polynucleotide. The polynucleotides and encoded  
proteins are useful as pharmaceutical agents and many disease-related  
genes may be included in them, for developing a diagnostic marker or  
medicines for regulation of their expression and activity, or as targets  
of gene therapy. The genes are involved in tissue and/or cell  
regeneration. Membrane proteins, signal transduction-related proteins,  
transcription-related proteins, disease-related proteins and genes  
encoding them can be used as indicators for diseases (e.g. osteoporosis,  
neurological diseases, cancer, tumours). The cDNA may be used to regulate  
the activity or expression of the encoded protein to treat diseases. The  
sequence presented is a protein of the invention. Note: Some of the  
sequence data for this patent is not represented in the printed  
specification, but is based on sequence information supplied by the  
European Patent Office.

XX Sequence 933 AA;  
SQ  
Query Match 25.8%; Score 2126; DB 7; Length 933;  
Best Local Similarity 49.6%; Pred. No. 9.7e-118;  
Matches 389; Conservative 117; Mismatches 239; Indels 40; Gaps 12;  
QY 2 VRCIQKLNRTVANEICEHFALQPTQACGLIPCRDVCVSEFLPWNCSKGGCKLQHR 61  
DB 158 IACIQKDKDIPAEIDIICEYFEPKLEQACGLIPQDCIVSEFSANSECKTCSGGLQHR 217  
QY 62 TRAVIAPPLFGGLQCPNLTSRACDAPISCPLEBEYTFSLKYGPWPKRLPHLKEINPS 121  
DB 218 TRHVVAPPQFGSGCPNLTFQVCQ---SSPCEAEELRYSLHYGPNSTCMSPHSRQVQA 274  
QY 122 GRT-----VLDNFNSDSNERVTEK-HQSYKAHHSSKSWAJEIGYQYQYQVSCTR 167  
DB 275 RRRGNKEREKDRSGKYVD--PEARLIIKKRNRNQRQENKYWDIQLQYQYREVWCIN 332  
QY 168 SDGNAMLSLCLQDSFPLTVQSCIMPDKCETSWMSWSPCKTCSRGSLLPGFRSRNV 227  
DB 333 KTGKAADLSFCQBEKLEPMTFQSCVITKECOVSEWSEWSPCKTCHDMVSPAGTRVTRTI 392  
QY 228 KHMAGGKCEPPELLEKEACIVEGELLQOCPRYSWRTSEWKECOVSLLEQODPHWVTG 287  
DB 393 RQFFIGSEKCEPEFEKEPCLSQDGVVPCATYGMRTTEWTECRVDPPLLSQDQKRRNQ 452  
QY 288 PVOGGGIQTRVYCAQS-----VFAAALRAKEYSRPVEKALCVGPAPLPQOLCNIPSTD 343  
DB 453 ALCGGGIOTREVYCVQANENLLSLSLTHKNKEASKPMDLKLTGPIPTNTQLCHIPCPT 512  
QY 344 CIVSSWSAWGLCTHENCHEPQKGGKGFTRORHVLMESTGPA---GHCPLHVESVPCEDPM 400  
DB 513 CEVSPWSAWGPCTYENCNDQGGKGFKLRRRTNEPTGGGVTGNCPLHLEAPCEEPA 572  
QY 401 CYRWLASE-GICFPDGHK-CGLGHRILKAVCONDRGEDVSGSLC-----PVPPPPERKSC 453  
DB 573 CYDWKAVRLGNCEPDNGKECGPTGVQVQEVVCIINSDEVDVQLCRDALFIP-----VAC 627  
QY 454 EIPCRMDCVLSEWTEWSSCSQSCSNKNSDKQTRSRITILALAG-EGKPPPPSQALQHR 512  
DB 628 DAPCPKDCVLSTWSTWSSCSHTCSGKTEGKQIRARSILAYAGEEGGIRCPNSSALQEV 687  
QY 513 LCNDHSCQMLHWETS PMGPCSEDTLVLTALNATIGWGEATCGVGIOTRRVFCVKSHVGOV 572  
DB 688 SCNEHPTVYHWGTGPGWGCIEDTSVSSFNITTTWGEASCSCVGMQTRKVICRVNVGV 747  
QY 573 MTKRCPDSTRPVRPCFLPCKKDCIVTAFSEWTPCPRMQAGNATV-KQSRVRIIIQEA 631  
DB 748 GPKKCPESLRPETVRPCLLPCKKDCIVTPYSDWTSCPSSCKEGDSSIRKQSRHRVITQLP 807  
QY 632 ANGSOECPTLYERECEVDVSLCPVYRWKPKWSPCILVPESVWQGITGSSRACGKLOT 691  
DB 808 ANGRCCTDPLYEKACEAPQACQSYRWKTHKWRRCQLVWVSQQDSPGAQEGCGGROA 867  
QY 692 RAVSCISDDNRSAEMMECLKQTNMPELLVQECTVPCREDCTFTAWSKFTPCSTNCEATKS 751  
DB 868 RAITCRKQDGGQAGIHECLQYAGVFPALITQACQIPQDDQLTWSKFSKSSCNGDCGAVRT 927  
QY 752 RRRL 756  
DB 928 RKRTL 932  
RESULT 15  
ABR58460  
ID ABR58460 standard; protein; 577 AA.  
XX ABR58460;  
AC ABR58460;  
XX 07-JUL-2003 (first entry)  
XX Human NOV47k.  
DE



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 24, 2004, 00:32:36 ; Search time 51.0517 Seconds  
(without alignments)  
2760.353 Million cell updates/sec

itle: US-10-022-710-2

effect score: 8241  
equence: 1 MVRGIQLNRIVANICEH.....QSTPPQKPLTLAYDGLDM 1465

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 283366 seqs, 96191526 residues

otal number of hits satisfying chosen parameters: 283366

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	694	8.4	2165	2 T21371	hypothetical prote
2	587.5	7.1	1444	2 T18856	angiogenesis inhib
3	429	5.2	1059	2 T22545	hypothetical prote
4	428.5	5.2	807	2 A38152	F-spondin - rat
5	405.5	4.9	1558	2 C89114	protein G37C3.6a {
6	405.5	4.9	2167	2 T34395	hypothetical prote
7	393.5	4.8	803	2 A47723	F-spondin precursor
8	357.5	4.3	1700	2 S08167	Balbani ring 3 pr
9	345.5	4.2	1074	2 JC5928	senaphorin F precu
10	345	4.2	988	2 T14764	hypothetical prote
11	340.5	4.1	788	2 T25061	hypothetical prote
12	340.5	4.1	951	2 T00260	hypothetical prote
13	335	4.1	957	2 T15976	hypothetical prote
14	335	4.1	5376	2 T42215	zonadhesin - mouse
15	333	4.0	1360	2 T33922	hypothetical prote
16	316	3.8	805	2 T34212	hypothetical prote
17	312.5	3.8	469	1 S23126	properdin precursor
18	310.5	3.8	2321	2 S78549	notch3 protein - h
19	309	3.7	2318	2 S43306	notch 3 protein - bovi
20	305	3.7	2871	2 A55567	fibrillin 1 - bovi
21	302.5	3.7	1584	2 T00026	brain-specific ang
22	302	3.7	2703	1 A24420	notch protein - fir
23	298	3.6	2437	2 S42612	transmembrane prot
24	297	3.6	2807	2 A5278	fibrillin-2 precu
25	293.5	3.6	2555	2 A40043	notch protein homo
26	292.5	3.5	2518	2 A54105	fibrillin-2 precu
27	292	3.5	2531	2 A46019	notch-1 protein -
28	284	3.4	2531	2 S18188	notch protein homo
29	283	3.4	1620	2 T27283	hypothetical prote

fibrillin 1 precu  
MEGF6 protein - ra  
hypothetical prote  
immunodominant mic  
protein CTBP - mal  
properdin - mouse  
serine proteinase  
Notch homolog prot  
thrombospondin 1 p  
fibrillin-1 precu  
antigen Em100 - B1  
hypothetical prote  
Xotch protein - AF  
brain-specific ang  
hypothetical prote  
brain-specific ang

## ALIGNMENTS

### RESULT 1

T21371

hypothetical protein F25H8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T21371; T24896

R:Gajadsty, S.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19413

A:Accession: T21371

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-2165 <N1L>

A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3

A:Experimental source: clone F25H8

R:Gajadsty, S.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19949

A:Accession: T24896

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-2165 <W12>

A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3

A:Experimental source: clone T13H10

C:Genetics:

A:Gene: CESP:F25H8.3

A:Map position: 4

A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 814

Query Match 8.4%; Score 694; DB 2; Length 2165;

Best Local Similarity 22.9%; Pred.No. 3.7e-36;

Matches 351; Conservative 133; Mismatches 556; Indels 492; Gaps 84;

Qy 18 CEHALQPPTQACLPDPR-----DCVSEFLPWNCSKGGCKKLOHTRAVIAP-PLF 71

Db 586 CHH-----GACVRLAPESLTIKIDQMGDMWGECSRTCGGVQKGLRDCDSKPRN 637

Qy 72 GGLQCPNLTES-RACDAPISCPLEBEYTFSLKVGPSKRLPFLHKEINPNSGRTVLDPNS 130

Db 638 GGYCVGQRYRYSCTNQ-ECFWDTPQY-----REVQCS-----EFN- 673

Qy 131 DSNERTVFKHQYKAHHHSKSWAIEIGYQTRQVSCSTRDQGNMLSL-----CLQD 181

Db 674 --NKDIGIQGVASTNTHTWPKYANVAPNERCKLYC-RLSGSAAYLLLRKVDVGTCDEN 730

Qy 182 SFPLTVQSCIMPKDCE-----TSQWSSWSPCKTCSGSLIPG----- 219

Db 731 GDDICVAGACMPAGCDHQLHSTLRDKVCGVGGDDSCKVVKGTFTNEQGTFGVNEWKIP 790

Qy 220 FRSRNRNVKMAIGGGKCEKPELLKEACIVEGELLQCCPRYSWRTSEWKECQVSLLEQ- 278

Db 791 AGSANIDIRKQYNNKEDDNYLSLA--ANGEFLLN-----GHFQVSLARQOI 837

QY 279 --OD-----BHWHTGP-----VCGGGIQTREV---YCAQSVPAALARA 313  
Db 838 AFQDTVLEYSGDALIERINGTPIRSDIYVHLVSGSHPPDISYEWTAAPVNA-----892  
QY 314 KEVSRPVEKAL-----CVGAP-----LPSQ- 334  
Db 893 --VIRPISSALYLWRVTDWTTECDRACRQGSQKLMCLDMSTHQSHDRNCQVLPKQA 950  
QY 335 --LCNIPCSDCIVSSWSAWGLCIHENCHEPOGKGRTRQHVLMESTGPAGCHPLVE 392  
Db 951 TRMCNIDCSTRWITEDVS-----SCSAKCSGQKQKQVSCVKEGEBQRTFASHLCD 1002  
QY 393 --SVPCBDPMCY-----RWLASGI-CFDPHGKGLGHRILKAVONDEGEDVSGSLC- 442  
Db 1003 RNSKPSDIASCIYDCSGKKNYGEWTSCTSGNSGKWHR--KSYCVDDSNRRVDESCLG 1060  
QY 443 -PVPPPERKSCIEPCRMDCVLSWMT--EWSSCQSCKNSKNSDKQTRSRITILALAGE-- 497  
Db 1061 REQEATERECNRIPC-----PRWVGHSECSRCDG-----GVKRWHAQCLDAADRET 1110  
QY 498 GKGPCPPSQALQEHRLCNDSHSMOLHWETSPWGPCESEDTLVTALNATIGWNGEATCGVGI 557  
Db 1111 HTRSCGPAQT-QEH--CNEHAC--TWQFGVWSDCS-----AKCGDGV 1148  
QY 558 QTRRVFCVSHVGQWTKRCPDSTRPSTVRPCFLPCKKDCIVTAFSEWTPCPRMCO-----613  
Db 1149 QYRDANCTDRHRSVLPBHRCLKMEKIIT-KPCH---RESCPKYKLGWSSQCSVSCDGS 1204  
QY 614 -----AGNAT-----VKQSRVRIIIQBAANGQCEP-----639  
Db 1205 SRRVSCVSGNGTEVMSLGTASDRPASHQTCNLG--TCPFRWNTDKSACSVSCGIGHRR 1263  
QY 640 --DTLYEREBEDVSLC-----PVRWKPKQWSPCILVPRESVWQGTGSS 682  
Db 1264 TTECIYREQSV-DASFCGDTKVPETSQTCHELLPCTSKWPSHWSPC-----S 1308  
QY 683 EACGKGQTRAVSC-----ISDDNRSAEMWE-CLKQINGMPLLVQECTVPCRE 729  
Db 1309 VTCSGGTQTRSVSCTRSGEIVDEYFCDNTRERLAKTCEKOTCDGPRVLQK-----LOA 1364  
QY 730 DCTTTAMS--KFTPCSNCEATKRRRQLTGKSRKKEKQD-----SDLYPLVET 777  
Db 1365 DVPTIRWATGPWTACATCG-----NGTQRLLKCRDHVADLDPDECNHLDEKST 1415  
QY 778 ELCPDCEFIQPYGNWSCILPEGRRRPHRLRLRYQADSKSCGELRFRVACSDXNGRPV 837  
Db 1416 RNCRLRDCSYWMAEWECEPATCG--THVQQRNVTCSAEDGGRTILKDVDC-DVQKRPT 1473  
QY 838 DPFCSSSGYIQEKCVIPCP--FDCKLSDW--SSWGSCTSSCGIGVIRIRSKWLKEKPYNGG 894  
Db 1474 SARNCR-----LEPCPKGEEHIGSWIIGDWSKCSASCGGWRRRSVSCTSSCDET 1524  
QY 895 RPKCLDLKNQVHEAVPCYSECQ-----YSWVVEHSSCKINNELSLRCGGGTQS 946  
Db 1525 RK-PKQ-----FDKCNELCPPLTNNSWQISPWTHC-----SVSCGGVQR 1564  
QY 947 RKIRCNTADGEGGAVDNLQNDQEI-----PPEQTSC-----SL 981  
Db 1565 RKINCEDVLSGR-----KQDDIECSEIKPREQDCMPPCPSHYHNKTSASMTSL 1615  
QY 982 MCPNECVMS-----EW--GLWSKCPQSCDPHTMQRRTHLLRPSLN-----S 1021  
Db 1616 SSSNSNTTSSASASLPILPFWSNQTSAMACSAKCGRT--KERVVECVNPSLNTVTAS 1674  
QY 1022 RTCAEDSQVPC-----LLNENCFOYNLTWSTCOLSENAPCGQVTRLLSCVCSGDK 1077  
Db 1675 TEC--DQTKPVEVRCRKIC--PRWKTITWSSCVT-----CGGIRREVQCYGRKN 1726  
QY 1078 PVSMDQCEQHNLEKQPMRSIPLVECV--VNCOLSGW--TAWTECSQTCGHRGMRSTRFI 1134  
Db 1727 LVSDSECN-----PKTKLANSVANCFFVACPAYRWNVTPWSKCKBCARGQKQTRVHC 1779

QY 1135 IMPTQGEGRPCPTBELTOEKT-----CPVTPC--YSWVLGNWSACKLEGDCGEGVQIRSLSC 1189  
Db 1780 ISTSGRAAPRMCELARAPTSIRSCDTSNCPYEWVPGDWQTC-----SKSCGEGVQTRVRC 1836  
QY 1190 --MVHSGISHAARVEDALCEMPFQDSILKQLCSV-PCPGD---CHLT-----EWS- 1236  
Db 1837 RKKNFNSTPIIFWLED-----BPAVPEKCELFPPKNESQTCELNAPCDSEFKWSF 1888  
QY 1237 -EWSCTCELTCT-----IDGRSFETV 1254  
Db 1889 GPWGECSNCGGIRRRRVKCVANDGRRVERV 1920  
RESULT 2  
T18856  
angio genesis inhibitor homolog - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T18856; T24653  
R:McMurray, A.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: Z19031  
A:Accession: T18856  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1444 <W1>  
A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:CO2B4.1  
A:Experimental source: clone C02B4  
R:McMurray, A.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: Z19917  
A:Accession: T24653  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1444 <W12>  
A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:CO2B4.1  
A:Experimental source: clone T07C5  
C:Genetics:  
A:Gene: CESP:CO2B4.1  
A:Map position: X  
A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3; 566  
Query Match 7.1%; Score 587.5; DB 2; Length 1444;  
Best Local Similarity 23.4%; Pred. No. 1.5e-29;  
Matches 315; Conservative 155; Mismatches 460; Indels 419; Gaps 86;  
QY 12 VVANEI-----CEHFALQPTTEQAQLPCPRDCVVSF-----LPWSNCS-----51  
Db 368 VLAHEMGNHMGVHDGVQVQCNKGC-----CLMSAVNGAGKTTWSDCSVREFNALILQ 420  
QY 52 ---KCGCKKLQHRTRAVIAPPLFGGLQCP--NLTESACDAPISCPLEBEETPSLVKVP 106  
Db 421 LDESGRGNCLRDASPLISTNHLSDRLPGQRTADQCC-----SYFWGR 465  
QY 107 WSKRLPHLKEINPSGRVLDNFSDSNERNVTFKQSYAHHSKSWALEIGYQTRVQSVCT 166  
Db 466 DYKVEIPNGKAMDDICRILWCGNSGST--ISTAHPALEG-----SWC-----505  
QY 167 RSDGQNAL--SLCLQDSFPPLTVQSCIMPKDCETSOWSSWSPSCSKTCR-----SGSLLP 218  
Db 506 ---GANKWCHKQCTHTWFTGLTP-----VPID--GESEWGAEGKGCPIQCCAVSGSITV 555  
QY 219 GPRSRNRVKNHAIIGGKECPLELKEAC-----IVEGELLQCCPRYSWRTSEWKECQVSL 274  
Db 556 QGCHRD--CVNPAPNNGKTC-----EGANIRGIVCGATSSNCLGFTREEFGNKICS--SI 607  
QY 275 LLEQQDQPHWHVTGPVCGGGIOTREYVCAQSVPAALRAKEVSRPVERKALCVGPAPLPSQ 334  
Db 608 KYDPHPQDQLTGEGEFHSTQPCRVWC--HLIGSLIRNKG-----646  
QY 335 LCNIPCSDCIVSSWSAWGLCIHENCHEPOGKGRTRQHVLMESTGPAGCHPLVESV 394  
Db 647 --QFPDGTFCGPDAYCVGGQCLALSC-----DNKALVEQ---PDCCPRI-----685



```
1091 -----KPO-----RMSIPCLVCVNVNQLSGWTAMTECSQTGCGG 1126
1127 RMGRTRFIIMPTQEG-----RCPPTLTQKTCPTVPCYSWVLGNWSACKLEGGDCG-B 1180
795 NVSLIEALLTAPNDEKAREQLRKYGNELVAR-----WDIGHWSECRQK--TCHVA 842
1181 GVOIRSLSCMV-HSGSI-SHAAGRVEDALCGEMPFQDSILKQLCSV-----PC-PGDCHL 1232
843 GYQARGISCKVTFHGEIRN-----VDNSIC-----ESLASVRPPTPRCHREDCPR 888
1233 TEWSEWSTC-ELTCDIGRSPETVGRSRTFTIIQSFENQDSQPQ--VLETRPCTGGKC 1289
889 WEASOWSECSQRCSV-----SMLAQKRNVTCTFTNGTSVDIQHCDITNRPATTMD 941
1290 YHYT-----WKASLWNNRITVWCORSQGVNVTG-----CSQPAR 1325
942 PNQCKAEWRTSDWGS-----CSSECG-----TGGVQLRLSCVWISSGRPAGRNCQMR 992
1326 PAATROCI-----PACRKPFSYCTQ-GGVCGCEKGYTEIMK 1360
993 PHSARACVADEPLPCMPPTASALYQRDASQDOSRFDIHK 1033

RESULT 4
A38152
C:Species: Rattus norvegicus (Norway rat)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 17-Nov-2000
C:Accession: A38152
R:Klar, A.; Baldassare, M.; Jessell, T.M.
Cell 69, 95-110, 1992
A:Title: F-spondin: a gene expressed at high levels in the floor plate encodes a secreted
A:Reference number: A38152; MUID:92208952; PMID:1555244
A:Accession: A38152
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-807 <KLA>
A:Cross-references: GB:M88469; NID:g204176; PID:AAA1174.1; PID:g204177
A:Experimental source: embryo floor plate
C:Superfamily: F-spondin; thrombospondin type 1 repeat homology
F:441-495/Domain: thrombospondin type 1 repeat homology <THR2>
F:500-555/Domain: thrombospondin type 1 repeat homology <THR3>
F:557-611/Domain: thrombospondin type 1 repeat homology <THR4>
F:613-666/Domain: thrombospondin type 1 repeat homology <THR1>
F:667-721/Domain: thrombospondin type 1 repeat homology <THR5>
F:753-807/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 5.2%; Score 428.5; DB 2; Length 807;
Best Local Similarity 28.8%; Pred. No. 1.2e-19;
Matches 133; Conservative 64; Mismatches 158; Indels 107; Gaps 27;

849 QEKVIRPCPFCKLSDWSSWGSCSS--CGIGVRIRSKWLKEKPYNGRPPCLDLKNQVH 907
432 EEKEDDTFETCIYNNSPWSACSSTCEKGRVQRMLKAQ-LDLVPCP-----DTQ 484
908 EAVPCY-----SECNQYSVWVEHWSCKINNELRSRLCGGTSQRKIRCVNTADGE 958
485 DFQPCMGFGCDEDEGSTCTMSEWIT--WSPC-----SVSCGNGMRSRE-RYXKOPPED 534
959 GGAVDSNLGNQDEIP-PEITQCSL--MCPNECVSEWGLWSKCPQCDPHTMQRTRHL 1014
535 G-----SVC-----MLPTEBTEKCTVNECSFSSCLVTEWGEWDGCSATCGMKGRHVX 586
1015 LRPSLNRITCAEDSQVQCLLNENCFQVNL-----TEWSTCQLSENAPCGQVRT--RL 1069
587 MSPADGSMCKAETSAQKCMWPE-CHTIPCLLSPWSESDCSVT-----CGKMRTRQML 641
1070 SCVCSDDGKPVMDQEOHNLKPPORMSIPCLVECVNVCOLSGWTATWTECSOTCGHGRMS 1129
642 KSLAELG-----DQNE-DLEQAEKMLP---ECPIQCELSWSEWSECNKSCOK-GHMI 690
```

```
1130 RTRFIIMTQEGRPPCTELTQKTCPTVPCYSWVLGNWSACKLEGGDCGEGVQIRSLSC 1189
691 RTRIIQMEPQFGAPCP-ETVQRKKCRARKC-----LRSPSIQKLWRPEAR- 736
1190 MVHSGSISHAAGRVEDALCGEMPFQDSILKQLCSVPCPGDCHLTWSEWSTCBLTCDG- 1248
737 -----SRSEQURESDGEQ-----PPG-CRMRPMTAWSECTKLCGGGI 774
1249 -RSPETVGRSRTFTIIQSFENQDSQPQVLETRPCTGGKC 1289
775 QERYNTVKKRFPKSSQF-----TSCKDK-KEIRACNVHPC 807

RESULT 5
C89114
Protein C37C3.6a [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C89114
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A750000; MUID:99069613; PMID:9851915
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: C89114
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1558 <SFO>
A:Cross-references: GB:chr_V; PIDN:AAC25867.1; PID:g3294501; GSPDB:GN00023; CBSP:C37C3.6
C:Genetics:
A:Gene: C37C3.6a
A:Map position: 5

Query Match 4.9%; Score 405.5; DB 2; Length 1558;
Best Local Similarity 20.3%; Pred. No. 6.6e-18;
Matches 322; Conservative 136; Mismatches 502; Indels 629; Gaps 86;

46 PW---SNGSKGCKKLQHRTRAVIAPLFGGLQCPNLTESACDAP-----ISCPLG 94
81 PWPVNECSRSCG-----GGVQ-----LEKQCSGDCGTGASVRYISCNLN 120
95 EEEYFSLKVGPKWCKRPLHLKEINPSGRTVLDNSDSNERVTFKHQSYKAHHSK--- 150
121 ACESGTDFAECCK-----FNDALDGNHYHKWTFYKGNKCELVKC 162
151 -----SWAIEIGYQTRQVYSTRSDGQNAWLSLQDSPLTVQSCIMPDKCETQWS 202
163 PDSGNFYFKWADKV-----VDGTKCDSK-----SNDICVDGECPLFVGCDGKLGS 206
203 S--WSPCKTCSGSLLPGRFSR--SRNVKHAIGGKCEPPELLEKACIIVEGELLQCP 258
207 SUKFKCKCKDGDGCTKTIEGRFDENLS-----PGVHDIILKPEGATNI----- 252
259 RYSMTSEWKECQVSLLEQDDPHWHVTGPGVGGGIQOTREYVCAQSVPAALRAKEYSR 318
253 ---KIQEARSTNNALKNQSDHFLYNG---NGLIQ----- 282
319 PYEKALCVG-----PAPLPSQLCNIPCSITDCIVSSWSANGLCIHENCHPEQKKG 368
283 -YEKEVEYGGTIFVYDDAEPETLSAQ---GFLSBLTVA-----LIFRKGSRDTAIKYE 332
369 FRTRQRHVLMESTGAGCHPLHVESVPCEDPMCV-----RWLASEGICFPDHGKGLGHR 423
333 F-----SIPLLEEVDYMYKFDNW-----PCSVSCGKGQV 362
424 ILKAVQONDRQED-VSGSLP---VPPPPERKSEIPCRMDCVLSEW--TEWSSCSQSCS 477
363 TRNLYCIDGKNGKRVEDDLCEENNAKPEFKSCB---TVDCE-AEWFTGDWESCSSTCG 418
478 KNNSGKQTR-----SRTILALAGEGCKPCPPSQALQEHRLCNHDSCHMQLHWETS 527
```







1;Residues: 1-1700 <PAU>  
2;Cross-references: GB:X52263; NID:g7057; PIDN:CAA36506.1; PID:g7058

3;Genetics:  
4;Gene: B3

5;Map position: 4

6;Supfamily: unassigned Balbiani ring proteins

Query Match 4.3%; Score 357.5; DB 2; Length 1700;  
Best Local Similarity 19.6%; Pred. No. 8.3e-15;  
Matches 303; Conservative 146; Mismatches 595; Indels 499; Gaps 89;

2y 28 EQACLIPCR--DCVSEFLPKNSCKGKGLKHRTAVIAPPLFGGLQCINLTESRAC 85  
3b 127 EKSCACVFNADKCTAPQVNNKDTCCCGFVNQEPADGCTKPLIWDKVC-----RC 179  
4y 86 DAPISCPLEEYTFSLKVGPMW-----DWSDDSCSCCKGDKGCKGKIWKNNCRICFP--TAEP 228  
5b 180 ECPKXKDGKGR-----DWSDDSCSCCKGDKGCKGKIWKNNCRICFP--TAEP 228  
6y 121 SGRVLDNFSDNE-----RVTFKHQSYKAHHHSKW-----AIEIGVQTRQVSCTRSDQON 172  
7b 229 AGCSAPLKWDDDKCSCACPAKMEKKEKCVESGKIWNPTCECG--CAQLNCPDNKKAN 286  
8y 173 AMLSCLQLDSFPLTVQSCIMPXDCETSSWSSWSPCS-----KTCRSGSLLPQFRSR 223  
9b 287 KETQC-----EKEVKKNGGVFCXKSCVCPGGDKDKTCTAPQVYDG-----332  
10y 224 SRNVKMAIGGKECPPELLEKEACIVEGELLQOCPR--YSWRTSEWK--EQVSLLEQDDP 281  
11b 333 -----VACSCSPVNMQKA-----DQCPRFQWDEKCEKCEPVK-----368  
12y 282 H-----NHVT-----GVCVGGIOTREVVCAQSVPAALRAKEVSRPVEKALC 325  
13b 369 HDCKNGKWNDEITCQICPRDAPVCTAGKER-----CGES-----CECKINREPKG-C 417  
14y 326 VGPAPLPOLNIPCTDCIVSSWSAWG-----LCIHENCHEPOGKGFRTORHVLME 379  
15b 418 AKPLVNNENTCKVCVPADKQMSPGGSGSGSKFNKLTQCECDQASAKGLKWN-----471  
16y 380 STGPAGCHPLVESVCEDEPMCVRWLASGIC-----FPDHGKGLGHRILKAVQNDRG 434  
17b 472 -----ADTCCKE-----QPGMPPEG-CGKQTWISDKCKECSPTI-----TCQAPQI 513  
18y 435 EDVS--GSLCPVPPPPERKSCIEPCRMDCVLSEWTSWSSCSQSCSN--KNSDGKOT-RSR 489  
19b 514 LDINTCEKCPVNLQAQKEKCKSP-----RQWTD-SKLCCESTTPATCGKGTWGB 565  
20y 490 TILALAGEGKPCPPSQALQ-----EHLRNDHSCMQLHWETSP--WG-----PCSEDTL 537  
21b 566 ACQCICPGDKRNGKNGKFFDKPSCECKCKNNPTC-----TSPQVWDADDCECKCPDKQ 619  
22y 538 VTALNATIG--WNGEATCGVIGIOTRVRVFCVSHVGQVMTKRCPDSTRPRTVPCFLPC--593  
23b 620 KPOGGCDGGQKMN--DRVSCGCPVPAPDCTNGQIYNNI-----CAGCGOI 664  
24y 594 -KDCIVTAFSEPTPCRCQAGNATVKOSRYRIIIEAANGQBCPDPLYBERECEDVS 652  
25b 665 DKPSCPKQIYNNKTCDCCEPNG-----MKEPVGG--CGAKTWLDDCQ---706  
26y 653 LCPVYRWKQKNSPCLLVESVWQGITGSEACGKLQTRAV-----694  
27b 707 -----CDVCPKPGKGGTGAQKCDKTCCKCKEMPTGCGENKKNKWDDET 752  
28y 695 -----SCISDDNRSAMMECL-----KQTNGMPLLVQECTVPCREDCTPTAWSKTP 741  
29b 753 CDCVCPQKNTCIAPKWDATKSCICVNPFPKNSPQVLKDTCCCGQNVKSKAPQKRIE 812  
30y 742 CSTNCEATSKRRRLTGKSRKEKQDSDLYPLVETELPCDEFTISQPYGNWSDCILPEG 801  
31b 813 NITCDACP-----NKKQCK-----APLVWS-----DEF-----CDCVCPN- 842  
32y 802 RREPHRGLRVQADSKCEBGLRFAVACS--DKNGRPVDPSPFC--SSSGYIOEKCVIPCP--857

843 -----SASMKITCLSPKWNKVTCTCDCN--PPKPDCCPGTQKWMDDKCKGCPNA 890  
858 -FDC-----KLSDSSWGSCSSCGIG---VRIKSKLKEKPYNG-----GRPCPKLDLKNQ 905  
891 QTDGAGKKFNDF-----TSCGCPSPGKLDCTGNTKWSAETCTCGGDVNRNG--NLKNF 944  
906 VHEAVPCYSECNQYVYVHEHWSKNNELRSIRGCGGTOSRKIRCVNTADGEGGAVDSN 965  
945 NDNL--CQCECKN-----KQEMANCKSPRTWYDTC-----KVCCKN-----ADSD 984  
966 LCNQDEIPPTQSCSLMCPNECVNSEWGLWSKCFQ-----SCD-----PHTM 1007  
985 DCPKQPLWDDQ--CKCCPASAQNT-----CPANKRFIEKSCCECKSPMPSPIQGX 1036  
1008 QRRTRHLRLSPLSRSTCAE-----DSQVQPCLE---LNENCFQFYNLTEWSTQQLSENAP 1059  
1037 KWNEDKCVVECANVKTCEGQFQWCDNCK--CICPQVNTKCDKQKPIESKCEGCEQTQ 1095  
1060 CGQVTRILL--SCVSD-----GKPV---SMDQEQHNLEK-----QRMSIPLVECVN 1106  
1096 CKDGRFNSNLECGCLCDKPKGQVFDKNTCQCKCPN--QKPGDTCNGKDFCPLDCSK 1154  
1107 CQ-----LSGMTAMTECSTQTCGHGRMSRTRFIIMPTQEGRPCTELTQKTCPTVP 1159  
1155 CKNPKPANGCTGYQVWNEKCCQ-----ECPKD-KPKKQCP---1189  
1160 CYSWVG--NWS-----ACKLEGCGEGVQVRSLSQVHSGSISHAAGRVEDALCGEM 1211  
1190 -----GGQDWNHQCQCGCTPAFTCSNNQKYSNVSC-----SCGCPGKPKNGCPGNQ 1238  
1212 PFQDSILKQLC--SVPCPGDCHLTSEWSTCLTICDG-----RSFET 1253  
1239 IWCNTRCVCPRMEXPADNCKTKWNDENCCVCKPGCPGEGCKGKWKNNANTSCCEC 1298  
1254 VGRSRRFTIIISFENQDSCPOQVLETRECTGCKVHYTWKASLWNNNERTYWCQRSDG 1313  
1299 PADKAKPASGDKKSWNDSCQCKSKMPC--GCGPPNQW-----NEKDCCKCS--1348  
1314 VNTGGC--SQAPAAIROC--IPA---CRKPFYCTGGVCGC 1351  
1349 --ATGNCFAGTWNSTQCSQSPATGKCTGAQVWCCKACKVC 1389

## RESULT 9

JC5928

semaphorin F precursor - human

C/Species: Homo sapiens (man)

C/Date: 10-Apr-1998 #sequence\_revision 08-May-1998 #text\_change 17-Nov-2000

C/Accession: JC5928

R/Simmons, A.D.; Poeschel, A.W.; McPherson, J.D.; Overhauser, J.; Lovett, M.

Biochem. Biophys. Res. Commun. 242, 685-691, 1998

A/Title: Molecular cloning and mapping of human semaphorin F from the Cri-du-chat candid

A/Reference number: JC5928; MUID:98125554; PMID:9464278

A/Accession: JC5928

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-1074 &lt;SIM&gt;

A/Cross-references: GB:U52840; NID:g2772583; PIDN:AAC09473.1; PID:g2772584

A/Experimental source: brain

C/Comment: This protein disrupts normal brain development and leads to some of the featu.

C/Genetics:

A/Gene: semaf

C/Superfamily: human semaphorin F; thrombospondin type 1 repeat homology

F/1-20/Domain: signal sequence #status predicted &lt;SIG&gt;

F/50-533/Domain: semaphorin #status predicted &lt;SEM&gt;

F/840-896/Domain: thrombospondin type 1 repeat homology &lt;THR3&gt;

F/971-993/Domain: transmembrane #status predicted &lt;TM&gt;

Query Match

Best Local Similarity 22.6%; Pred. No. 3e-14;

Matches 141; Conservative 59; Mismatches 168; Indels 255; Gaps 30;

794 SDCTL-----PEGRREPGLRVQADSKCEGGLAFRAV-----827

```
Db 441 SSCLEIEIEFPERRRPIRSLQILHSQSVLFVGLREHVVKIPLKRCQFYRTSTCIGAQ 500
Qy 828 -----ACSDKN-----GRPV 837
Db 501 DPYCGWDVWKCTSLBESLSMTQWEQISACAPPRNLTVDGHGVMSPTPCTHTDGSV 560
Qy 838 DPFSCSSGYIOEKVPCP-----FDCK-----LSDWSSWGCSSSSCGIGV 879
Db 561 GSCLCRT-----RSCDSPAPQCGWQCEGFCMEIANGSRNGTPTMTWSPCSTTCGIGF 615
Qy 880 RIRSKWLKE-KPYNGGRPCPKLDLKNQVHEAVPCYSECNOYSWVVEH-----WSSCK 930
Db 616 QVRGRCNPTFRGGRCVGVQNRERY-----CNEHLLCPFMFTWGPHERC- 665
Qy 931 INNELSLRLOGGTOQRKIRCVNTADGEGAVDSNLNQNDEIPEPTQCSL-WCPNECV 989
Db 666 -----TAQCGGIGQARRRICENGPDGAG-----CN-----VEYQSCNTNCPKELKT 707
Qy 990 SEWGLMSKCPQSCDPHTMQRETRH-----LL-----RPSLNSRTCAEDSQVOPCLLN 1036
Db 708 TPWTPWTPVNISDNDGHDYEQFRITKARLADPNLLEVGQRQIEMRYCSDG-TSGCSTD 766
Qy 1037 ENCFQF-----QYNL-----TEWSTCOLSENAPCGQGVTRLLSCVCSGK-P-VSM 1081
Db 767 GLSGDFLRAGRYSAHTVNGAWSMTWSQC-----SRDCSRGIRNR--KRVCNNPEPKYGG 820
Qy 1082 DQCCHNLEKFORMSIPCLVECVNQLSGHTANTQSCOTCGHGRMSRTRFIIMPTQGE 1141
Db 821 MPCLGPSLEYQECNTLPCFVGVWSC-----WSPWTKSATCG-GGHYMRTRCSNP- 871
Qy 1142 GRPCPTLTQKTPVTPCYSWMLGNWSACKLEGDCGEGVQIRSLSCMVHSGSISHAAG 1201
Db 872 -----APAY-----GDICLGL-----883
Qy 1202 RVEALCQEMFPQDSILKQLCSVPQDCHUTSEWSTBELTICIDGRSPETVGRQSR 1261
Db 884 HTEALCN-----TQPCPES--WSEWSDWSECEAS-----GVQVRAR 918
Qy 1262 TFIQSFENQDSCPQVLETRPC 1284
Db 919 QCIL-LFPMGSCSGNTTESRCP 940

RESULT 10
T14764
hypothetical protein DKFp434H204.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T14764
R:Wambutt, R.; Heubner, D.; Newes, H.W.; Cassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A:Reference number: Z18181
A:Accession: T14764
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-898 <WAM>
A:Cross-references: EMBL:AL110226
A:Experimental source: adult testis; clone DKFp434H204
C:Genetics:
A:Note: DKFp434H204.1

Query Match 4.2%; Score 345; DB 2; Length 898;
Best Local Similarity 20.1%; Pred. No. 2.7e-14;
Matches 228; Conservative 102; Mismatches 363; Indels 444; Gaps 58;
Qy 229 HMAIGGKECPLELLEKACIVEGELLQQCPYRWSRTSEWKECOVSLLELQQDPHHHTGP 288
Db 21 HREAGHDEVPP-----PVSWHYGPWTKCTVT-----48
Qy 289 VCGGIGTQREVYCAQSVFAAAALRAKEYSRPVEKALCVGPAPLPQSLCNIPCTDCIVSS 348
Db 49 -CGRGVQRQNVICLE-----RQAGPVDE-----70
```

RESULT 11  
T25061  
hypothetical protein T21B6.3 - Caenorhabditis elegans

```
Qy 349 WSAWGLCIHENCHBPQKKGRTRQRHVLVMESTGPAGCHILVESVPCEDPMCYRWLASE 408
Db 71 -----EHC-DPLGRPDQQRK-----C-----SEQPCPARWAGE 99
Qy 409 -GICFPDGHKGGLGHRILKAVQNDRGEDVSGSLCP-----VPPPERKSC--RIPCRMD 460
Db 100 WQLCSSSCGPGGLSRAV--LCIRSVGLDEQSALEPPACEHLPRPTTETPCNRHVPCP-- 155
Qy 461 CVLSEWT--EWSSCQSCSNKSNKQKQTRSTIILALAGEGKPCPPQSQALQEHRLNDHS 518
Db 156 ---ATWAVGNMSQCSVTC-----GEGTQRENVLIC-TNDTGVPCDEAQQPAVEVTCSLPL 205
Qy 519 CQLHMETSPGPCSEDTLVTALNATLGNWGEATCGVGIOIRRVF-----CVKSH----- 568
Db 206 C---RW-----PLG-----TLGPEG-----GGSSSHHELFPNEADFFPHILARPS 243
Qy 569 -----VGQVMTKRCPTSTRPETVRPCL-----PCKK-DCIV 599
Db 244 PASSPKPGTMGNAIEEAPELDLP--GPVFVDFFYDYNFINFHEDLSYGFSEEPDLDL 300
Qy 600 TAFSEWTPCPMCOAGNATVKQSRVRIIIQEAANGQCEPDTLYEERECEOVSLCPVVRW 659
Db 301 AGTGRTTPPHSRPAAPST-----GSPVATEPPAAKEGV-LGP---W 340
Qy 660 KPQKW-----SPCILVPESVWQGTGSSEACGK 687
Db 341 SPSPFSCAGRSPPPPSBCTPCNPLINFLPEEDTIFIGADLGLPSLSWPRVSTD----- 394
Qy 688 GLQTRAVS-----CISDDNRSAEMECLEKQTN-----GMPLLVQECTVPCRE 729
Db 395 GLQTPATPESQNDFFVQKDSQSLPPPWDRDRTNEVFKDDEBPFGKRGAPHL-----PPRP 448
Qy 730 DCTFTAWSKFTPCSNCEATKRRRQLTKSKRKKCCODSLYPLVETELCP---CDEFI 786
Db 449 SSTLPLSPVG--STHSFSPDVAELTGTIVANEPALGGLP-VDSBLNFTVGVASIL 505
Qy 787 SOPYGNWSDCILPEGREPHRGLRYQADSKCEGELRFRACVACSKNGRPVD-----PS 840
Db 506 PPPIAP-----LPE-----MKVRDSLSLEPGTP-SPPAPGPGSWDLQTVAWVGTFLEPT 551
Qy 841 FCSSSGYIOEKVIP-----CPFDCKLSDWSSWGCSSSCSGIGVIRSKWLKE 888
Db 552 TLTGLGHMPEPALNPGKQPESLSFEVPLSRLLSTPAWDSPANSHRPV---ETQPLAP 608
Qy 889 KPYNGRPPCKLDLKNQVHEAVPCYSECNOYSWVVEHWSCKINNELSLRCGGGTQSRK 948
Db 609 SLAEGAPPADPLVVERNA-----SWQAGNWSEC-----STTCGLGAVWRP 647
Qy 949 IRCVNTADGEGAVDSNLNQNDEIPEPTQCSLSLMCPNECVMSWGLMSKCPQSCDPHTMQ 1008
Db 648 VRCSGRDED-----CAPAGREFQPARRCHL---RPCATWHSQWNSKSKRS----- 689
Qy 1009 RRTREHLRPSLNSRTCAEDSQVOPCLLNENCFQOYNLTWSTCOLSENAPCGQGVTRTL 1068
Db 690 -----CGGSSVYRD 698
Qy 1069 LSCV-CSDGKPYVMQCEQHNLEKQPM--SIPCLVECVNQLSGW--TAWTCSQTC 1122
Db 699 VQCVDRDLRLRPLRPHCPGPAKPAHPRCPGAQPCCL-----SWYTSSWRRECSEAC 748
Qy 1123 GHGGRMSRTEFIIMPTQGRGRCPTLTQEK--CPVTPCYSWVLGNWSACKLEGDCGE 1180
Db 749 GGG---EQORLVTCPEFG---LCEERALRENTTRPCNTHCTQWVVGPGWQC---SAPCG 799
Qy 1181 GVQIRSLSCWHSVSGSISHAAGVED--ALCGMPFQDSILKQLCSVPC-PGDCHLITE 1234
Db 800 GVQRRLVKC-----VNTQTGLPEEDSDQCGHEAWPES-----SRPCGTEDCEPVE 844
```

Species: Caenorhabditis elegans  
Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
Accession: T25061  
Submitted to the EMBL Data Library, November 1995  
Reference number: Z19975  
Accession: T25061  
Status: preliminary; translated from GB/EMBL/DBJ  
Molecule type: DNA  
Residues: 1-788 <WIL>  
Cross-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3  
Experimental source: clone T21B6  
Genetics:  
Gene: CESP:T21B6.3  
Map position: X  
Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 4.1%; Score 340.5; DB 2; Length 788;  
Best Local Similarity 26.6%; Pred. No. 4.6e-14;  
Matches 120; Conservative 51; Mismatches 145; Indels 135; Gaps 25;

2Y 773 PLVETELCDFEFISQYGNWSD---CILPEGRREPHRLGRVQADSKEGEGRLRFAVAC 829  
DB 441 PACDGGQCVNPVVGWHDWSDWSTC-----SCTCGDGAQRREC 482

2Y 830 SDKNGRPVDFSCSSSGYIQEKVI-PCPFDCKLSDWSSWGSSCGIGVIRSKWLKE 888  
DB 483 STNN-----CQGADETEPCNLGFCO---TWSEWCWSTCSACSGGQGRTRFC-- 529

2Y 889 KPYNGRGPCKLDKNQVHEAVPCYSECNQYVWVHNSCKINNELRLSLRCGGGTQSRK 948  
DB 530 --HLGTRNCEKDYSEQCSAGPC-FEWSQW---EDWQCQ-----SVTCGGQVAVRQ 575

2Y 949 IRCVNTADGEGGAVDNLNQNQBIPETOS-----CSLMPCNECWSEWGLWKCQPS 1001  
DB 576 RTCL-----GGVFGHLQO---GPKTRACDGGFCFLWSP-----WQENWTCAS 618

2Y 1002 CDPHTMORTRHLLRPSLNSRTCAEDSQV---QPCU-LNENCFQFOYNLTWSTQLSN 1057  
DB 619 CG-SGNKRQR-VQFGTCCQGNESQFCYGPFCAEWTWC-----EWSGC-----S 664

2Y 1058 APGQGVTRLLSCVSDGKPVSMQCEQHNLEKFORMSIPCLVECVNQCQSGWTAMTE 1117  
DB 665 SKGPGQRTTRGCLGNQGEAT--TCQGPSIE-----TTLCEQSCCNWSEWCHWSM 715

2Y 1118 CSOTCGHGGMSRTREI--IMPQGEGRCPPELTQETKCPVTPCYSWVLGNWSACKLEG 1175  
DB 716 CDKECGG---QVRYEYFNFTGCEWSPCSTOLA----- 746

2Y 1176 GDCGEGVQIRSLSCMVHSGSISHAAGRVEDA 1206  
DB 747 --CEVGQSRROCVGESG--CHCIGLAES 773

RESULT 12  
T00260  
hypothetical protein KIAA0605 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
C:Accession: T00260  
C:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.  
DNA Res. 5, 31-39, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete  
A:Reference number: Z14086; MUID:98290545; PMID:9628581  
A:Accession: T00260  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-951 <NAG>  
A:Cross-references: EMBL:AB011177; NID:G3043733; PIDN:BAA25531.1; PID:G3043734  
A:Experimental source: brain  
C:Genetics:  
A:Note: KIAA0605  
C:Superfamily: thrombospondin type 1 repeat homology

F:46-106/Domain: thrombospondin type 1 repeat homology &lt;THR1&gt;

Query Match 4.1%; Score 340.5; DB 2; Length 951;  
Best Local Similarity 18.4%; Pred. No. 5.6e-14;  
Matches 219; Conservative 117; Mismatches 357; Indels 499; Gaps 58;

QY 200 QWSWSFCSKTCRSGSL-----PGFRSRNVKMAIGGKSCP---EL 241  
DB 52 EWTKWTFASRSCGGVTSQBRHCLQRRKSVPGNRTCTGTSKYQLQCRVQCPDGRS 111

QY 242 LEKACIVEGELLQCCPRYSWRSEWKECCQVLSLLLEQDPMHVTGVCGGGQTRREVYC 301  
DB 112 PREQCYSFNSHV-----YNGRTHQWKL-----YPDYVHLSKPC-----DLHC 152

QY 302 AQ-----SVPAALRAKEVRPEKALCVGAPLPQLCNIPCTDTCIVSWSAWGL 354  
DB 153 TTVDGQRQLMVPARD-----GTSCKLTLDR--GV 179

QY 355 CIHENCHEPCKKG--PRTRQRHVLMESTGPAHCPH-----LVESVPC- 396  
DB 180 CVSGKC-EPICDGVLFSTHTLQKGCICQDGGSCHTVTGNRKGNHGLYSLVTHIPAG 238

QY 397 --EDPMCVRWLASEGICFPDGH-----KQGLG 421  
DB 239 ARDIQIVERKKSADVLALADEAGYFFENGNYKVDSPXNFNIAGTVKVRPMDVIETGIE 298

QY 422 HRLKAVC-----QNDRGEDVSGSLCPVPPPERKSCBPCBMDCVLSEWSS 471  
DB 299 YIVAGGTNGLNVMVWQNGKSPSITFYTLQPPHESRPQI-----YYG 345

QY 472 CSQCSKNKSDGK-----QTSRTI--LALAGEG--KPCPPQALQEH 512  
DB 346 FSESAGSLDGLAGLMGFPHNGSLYQASSERLGLDNLFGHPGLDMELGPGQGETNE 405

QY 513 LCND---HSCMLHWETSPWGPCSEDLTALNATIGWNGEAT----- 552  
DB 406 VCEQAGGAC-----EGPRGKGRDRNVGTGTLTGDKDDEVDTHFASQEFFSANAISD 460

QY 553 --CGVGIQTRVRFVKSHVQVMTKRCFDPSTREP-----VRPCLPCKK 595  
DB 461 QLLGAGDLKD--FTLNFTVNSIFAQAPRSLAESFVDEENEGAGPYLLNSYLESS 519

QY 596 DCIVTAFSEMTPCPR-----MCOAGNATVKQSYRIIOEAANGGOCPTLYEERECED 650  
DB 520 DRVANSSE-APPNVSTSLTSAGNRTHK--ARTR-----PKARKQG 559

QY 651 VSLCPVYRWKQWSPCLLVPSVWQGITGSSACGKGLQTRAVSCISDDNRSAEWMECL 710  
DB 560 VSPADMYRWKLSSEPC-----SATCTTGVMSAYAMCVRYDGVVD----- 600

QY 711 KQTNMPLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKSRRLQTKSRKKEKQDSD 770  
DB 601 -----DSYCD-----LTPRPVHEFCAGRE 621

QY 771 LYPLVETELCDFEFISQYGNWSDCILPEGRREPHRLGRVQADSKEGEGRLRFAVACS 830  
DB 622 CQPRWET-----SSWSEC-----SRTCGEGYQFRVVR- 649

QY 831 DKNRGPVDFSCSSGYIQEKVIPCPCFCDKCLSDWSSWGSSCGIGVIRSKWLKEKP 890  
DB 650 --WKMISGPF-DSVV-----SLCEAAEAVRE----- 675

QY 891 YNGRGPCKLDKNQVHEAVPCYSECNQYVWVHNSCKINNELRLSLRCGGGQTSRKIR 950  
DB 676 -----ERKTCRNPAQGPQWSEWSEC-----TAKGCE-----R 704

QY 951 CVNTADGEGGAVDNLNQNQBIPETQSCSLMPCNECWSEWGLWSKCPQSCDPHTMQR 1010  
DB 705 SVWTRDIR-----CSEDE-----KLCDDNT----- 724

QY 1011 TRHLLRPSLNSRTCAEDSQVQPCILLNENCFQFOYNLTWSTQLSNAPCGQGVTRFLLS 1070  
DB 725 -----RP-VGEKNCTG-----PPC-----DRQWTVSDWGPC-----SGSCGQGRTRHVV 763

QY 1071 CVCSDGKPVMSDQCEQHNLEKQPMRSIPCLVECVYNQOLSGWTA--WTEGQTCGHHGGM 1128  
Db 764 CKTSBGKVPESQOMET--KPLAIH-PCGDK--NCP-AHMLAQDWERCNTTCGRG-- 813  
QY 1129 SRTFIIIMPQGEGRP-----C-----PTELQTEKTCPTVCYSNVLGNWSACKLEGG 1176  
Db 814 VKRLVLICMELANGKQTRSGPECGGLAKKPE---ESTCFERPCFRWYTSFWSBECT--K 867  
QY 1177 DCGEGVOIRSLSCMVHSGSI-SHAAGRVEDALCGMPFQDSILKQLCSV-PCP 1227  
Db 868 TCGVGVRMDVKC--YQGT-----DIVRGCDPLVKPVGRQACDLQPCP 908  
RESULT 13  
T15976  
hypothetical protein F08C6.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15976  
R:Bentley, D.  
submitted to the EMBL Data Library, June 1995  
A:Description: The sequence of C. elegans cosmid F08C6.  
A:Reference number: Z18440  
A:Accession: T15976  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-957 <BEN>  
A:Cross-references: EMBL:U29378; NID:g868184; PID:g868185; PIDN:AAA68721.1; CESP:F08C6.1  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:F08C6.1  
A:Introns: 23/1; 135/3; 220/3; 285/3; 325/3; 376/3; 461/2; 498/3; 532/3; 627/2; 714/1; 7  
Query Match 4.1%; Score 335; DB 2; Length 957;  
Best Local Similarity 22.6%; Pred. No. 1.3e-13;  
Matches 169; Conservative 67; Mismatches 279; Indels 232; Gaps 42;  
QY 398 DPMCYRW---LASGICFPDH--KQCGLGHRLKAVQNDRGEDV--SGSLC----- 442  
Db 344 DGMCDPWNCTLASGLDFTSAFIGTHLGHFTL---KDKHSDTLFTGFCFQSKWCQL 400  
QY 443 -----PVPPPPERKS---CEIPCRMDCVLSEWESS--CSQSCSN--KNSDGKQ 495  
Db 401 GRCVPWTGTNEIQTVQHVAPVWVTLPSRID---GSWSGNGATICSQCTCNGILGVSGLA 457  
QY 486 TRSRITLA-LAGEGKPC--PPSQALQEHRLCNDHS-----CMQLHWETSPWGPC 532  
Db 458 IARITCAPYPANGSGDCVGSSTSAVLCSRCQGRASKSVDEYISDKMEQK-----RL 510  
QY 533 SEDTLVLTALNATIWNAGEATCGVGIQTRRVFC-VKSHVGQVMYTKR-----CPDSTRPETV 586  
Db 511 KNDRELTKGSQLNRFPQAC-----KVPDVOQHYGSRNRYRFGDNLDPDGTSCGYD 563  
QY 587 RPPC-----LPCKKDCIVT-----APSEWTPCPRCQAG 615  
Db 564 RYLDGELALNCNNALISRDQSCPTDTCPIITQSSSVRGQWGTWSLWTSCTATCGG 623  
QY 616 NATYKQRYRIIIIEAANGGQCEPTDLYERECECDVSLCPVYRWKPKQKSPCILVPESVW 675  
Db 624 -----YRKNRACSLTQ--CEGNEDETEVCSSS--CPSVLRVGNEM-----STW 665  
QY 676 QGITGSSEACKGLOTRAVSCISDDNESAEMEC---LKQTNGNPLLVQSCVTPCREDC 732  
Db 666 TEWNHCSVSGRGSGQARYKCKLSPHRTLA--FDCPGENKVTNELRI-----T 710  
QY 733 FTAWSKTTPCTNCEATKSRRRRLTGSRKKEKQDSLDLYPLVETELCPQDEF-ISOY 791  
Db 711 FFKARSYIMCVR--NKIKENTISEXNIEVRSNDG-----PCNAIGVWGWTG 757  
QY 792 NWSOCILPEGRBBRHLRGLRVQADSKEGEGLRFRVAVACSKDKNGRPVDPFSCSSGYIQEK 851  
Db 758 GWSTC-----STSCGPGTLVRQRTC--NREP-----CDGSAHERRS 791

QY 852 C-VIPPPDKCLSDWSSWGSCSSCGIGVIRIRSKWLKEKPYNGRPPCFKLDKQ-VHEA 909  
Db 792 CNVATCONDGISLWLNWSDCSRVCXGKGLRSRSC-----FGSGCMGASSEQOFCNEQ 845  
QY 910 VPCYSECHQY-SWVVEHWSSCKINNEILSLRACGGTOSRKIRCVNTADGEGAVDSNLGN 968  
Db 846 ACASSANDWGTW--SGWSQC-----SVSCGAGVK-RRTRTCRTGNCPNYKESAICN 895  
QY 969 ODEIPPETQSCSLMCPNE-CVWSEWGLMSKQPCSDPHMTMORRTHLLRPLSLNSRTCAED 1027  
Db 896 DRD-----CENKNAAGGAGWSSCSETCGDGVKX----- 926  
QY 1028 SOVQPCLLNENCFQFQYNLTWSTCOL 1054  
Db 927 --VRKCYGSGNCDGQY---EKQYCNL 948  
RESULT 14  
T42215  
zonadhesin - mouse  
N:Alternate names: sperm-specific membrane protein  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
A:Accession: T42215  
R:Gao, Z.; Garbers, D.L.  
J. Biol. Chem. 273, 3415-3421, 1998  
A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane pro  
A:Reference number: Z22080; MUID:98123114; PMID:9452463  
A:Accession: T42215  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-5376 <GAO>  
A:Cross-references: EMBL:U97068; NID:g3327420; PID:g3327421; PIDN:AAC26680.1  
C:Genetics:  
A:Gene: Zan  
A:Map position: 5  
C:Function:  
A:Description: functions in multiple cell adhesion processes  
A:Note: found exclusively on the apical region of the sperm head  
C:Keywords: cell adhesion  
Query Match 4.1%; Score 335; DB 2; Length 5376;  
Best Local Similarity 20.3%; Pred. No. 7.3e-13;  
Matches 325; Conservative 154; Mismatches 523; Indels 596; Gaps 97;  
QY 18 CEHFALQPTTEQ---ACLIQPCRCVSEFLPWNCSKCGKQLQHRTRAVIAPPLFGL 74  
Db 2933 CATITLQCPAHSHTNCLPQPCPSCLDS-----GHC----- 2964  
QY 75 QCPNLTESRACDAPISCPG---EEY-TFSLKVGPMKCLPHLKEI-NPSGRTVLDFN 129  
Db 2965 -----EGSTTKAPASACQCGVCEPDYVVLNNKCVPRIECGCKDAQGVLPADKTWI--N 3016  
QY 130 SDSNERTVFKH--QSYKAHHHSKSWAIEICVQTRQVQSCSTSDGQ---NAMLSLCLQDSF 183  
Db 3017 RCTQSCCTCKGALQCKQFQCPSETYCKDI--EAGNSCTRISLQCPANSNFTSCLPSCQ 3074  
QY 184 PLTVQSCI-MPKDCETSQWSSWSPCK--TCRSGLLLPGRFSRNRVKNHMAIGGKCEPE 240  
Db 3075 P-----SCNNTDVHCEGSSPNTLSSCREGVCQSG-----Y 3105  
QY 241 LLEKACIVEGELLQCCPRYSWRPSEWKECQVSLLEQDDPHWHVTPVC-----GGG 293  
Db 3106 VLHNDKCI---LRNQC-----GCKDAQALPEGKT--WITSG--CTQSNCTGGA 3149  
QY 294 IQTR-----EYVC-----AQSVPAALAAKEVS-----RPV---EKALCVGPAP 330  
Db 3150 IQCNFQCPKLTCKYKDLKDGSSNCTNIPLQCPAHSRYTNCLPSCPPLCLDPEGLCEGTS 3209  
QY 331 -LPSQLCNIPSTDCIVSSWSAWGLCIHEN-----CHEPQKKGFRTR-----QRH 375  
Db 3210 KVPS-----TREGICQP-----GYLMHKNKCVLRIFCGCKXTQG--AFISADKTWISRG 3258



```

QY 979 CSLMCPNECVMSWGLSKCPQSDPH--TWQRTRHLLRPLSLNRITCAEDSQVQPCLLN 1036
Db      :      :      :      :      :      :      :      :      :
1001 C---CPETGI---WSDWTSTENQCRDYCGSGNQR-----TRICTSDADGCPQ--- 1043
QY 1037 ENCQFQVNLTEWSTCQLSENAPCGQGVRTLLSCV-----CSDGKPVSMDOCEOH----- 1087
Db      :      :      :      :      :      :      :      :      :
1044 ----CGPTTITE-----PCGTGV-----CYFRLSCCPGYTATVEG-NQHICGP 1082
QY 1088 ---NLEKFPQMSIPCLVECVVNCQLSG-WTAWTECSQTCGHGGRMSRTRFIIMPTQG--- 1140
Db      :      :      :      :      :      :      :      :      :
1083 LTTAVADPKLN-TCGVSCCPFSAGIWGEWVSVCNDTCGSCQGETRKRKCLSLQYGCAC 1141
QY 1141 EGRCPTELTOEKC-----PVTPCYS----- 1162
Db      :      :      :      :      :      :      :      :      :
1142 TGNATDTSVCASSVCLPRPTSCCTGFKOMVNITGRIFYCGPLPVVPAFNPETTCDDPEK 1201
QY 1163 -----WVLGNWSACKLEGCGGCGVQIRLSLSCMVHSGSISHAAGRVEDALCGEMPFODS 1216
Db      :      :      :      :      :      :      :      :      :
1202 TGLMNDW--GAWTTCSATCGGC--GTQTRSKT-----CASAPY--- 1235
QY 1217 ILKQLCSVPFGDCHLTWSEWSSTCELTCDIGRSFETVGRQSRRTFIIOSFENQDSCPQ 1276
Db      :      :      :      :      :      :      :      :      :
1236 -----GCPCTGD--LTETQ--SCAKQVC-----TTGAQCCAGKEVATGYDGAQYCOD 1278
QY 1277 QVLETRPCTGCKYHYTWKASLWNNNERTVWCQRSDGV--NVTGCGSPQARPAIROCIP 1334
Db      :      :      :      :      :      :      :      :      :
1279 NTPEV--CTG-----TW-----TEWATLEGAVCNDTCGNCG-----IIP 1310
QY 1335 ACRKPFYSYCTGGVCGCEKGYT 1356
Db      :      :      :      :      :      :      :      :      :
1311 TSR-----YCFPSG-CQCSGAYT 1327

```

Search completed: February 24, 2004, 01:08:25  
Job time : 67.0517 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

MM protein - protein search, using sw model

Run on: February 23, 2004, 20:35:55 ; Search time 41.0236 Seconds  
(without alignments)  
1859.483 Million cell updates/sec

Title: US-10-022-710-2

Effect score: 8241

Sequence: 1 MVRCIQLNRTIVANEICEH.....QSTPPQKPLTLAYDGLDM 1465.

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	664	8.1	1935	AT99 HUMAN	Q92n4 homo sapien
2	628.5	7.6	1906	AT20 MOUSE	P59511 mus musculus
3	622	7.5	1911	AT20 HUMAN	P59510 homo sapien
4	428.5	5.2	807	FSP0 RAT	P35446 rattus norv
5	393.5	4.8	803	FSP0 XENLA	P35447 xenopus lae
6	360	4.4	1593	AT12 HUMAN	P58397 homo sapien
7	358.5	4.4	1023	SM5B MOUSE	Q60519 mus musculus
8	357.5	4.3	1700	BAR3 CHITE	Q03376 chironomus
9	351.5	4.3	1077	SMSA MOUSE	Q62217 mus musculus
10	345.5	4.2	1074	SMSA HUMAN	Q13591 homo sapien
11	336.5	4.1	1093	SM5B HUMAN	Q9P283 homo sapien
12	335	4.1	5376	ZAN MOUSE	O88799 mus musculus
13	324	3.9	867	SSPO BOVIN	P98167 bos taurus
14	322	3.9	1095	AT17 HUMAN	Q8e56 homo sapien
15	318.5	3.9	1077	AT10 HUMAN	Q9h324 homo sapien
16	314.5	3.8	1210	AT19 MOUSE	P59509 mus musculus
17	312.5	3.8	469	PROP MOUSE	P27918 homo sapien
18	310.5	3.8	2321	NTC3 HUMAN	Q9um47 homo sapien
19	309	3.7	2318	NTC3 MOUSE	M1982 mus musculus
20	306	3.7	1224	AT16 HUMAN	Q8e57 homo sapien
21	305	3.7	2871	FBN1 BOVIN	P98133 bos taurus
22	304.5	3.7	1207	AT19 HUMAN	Q8e59 homo sapien
23	303	3.7	2703	NOTC DROME	P07207 drosophila
24	302.5	3.7	1584	BAIL HUMAN	O14514 homo sapien
25	301	3.7	2871	FBN1 PIG	Q9tv36 sus scrofa
26	300.5	3.6	1696	PK55 BRACL	Q9nj15 brachioosto
27	298	3.6	2437	NTC1 BRARE	P46530 brachydanio
28	297.5	3.6	2319	NTC3 RAT	Q9r172 rattus norv
29	297.5	3.6	2911	FBN2 HUMAN	P35556 homo sapien
30	297	3.6	2907	FBN2 MOUSE	Q61555 mus musculus
31	292	3.5	2531	NTC1 MOUSE	Q01705 mus musculus
32	288.5	3.5	1223	AT14 HUMAN	Q8wx58 homo sapien
33	285.5	3.5	470	PROF_CAVPO	Q64181 cavia porce

## RESULT 1

ID	AT99_HUMAN	STANDARD;	PRT;	1935 AA.
AC	Q9P2N4; Q9NR29;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).			
GN	ADAMTS9 OR KIAA1312.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
[1]				
RN	SEQUENCE FROM N.A. (ISOFORM 3).			
RC	TISSUE=Fetal;			
RX	MEDLINE=20396138; PubMed=10936055;			
RA	Clark M.E., Kelnner G.S., Turbeville L.A., Boyer A., Arden K.A.,			
RA	Maki R.A.;			
RT	"ADAMTS 9, a novel member of the ADAM-TS/Metallospodin gene family.";			
RL	Genomics 67:343-350(2000).			
[2]				
RN	SEQUENCE FROM N.A. (ISOFORM 1), AND FUNCTION.			
RP	MEDLINE=22513925; PubMed=12514189;			
RX	Somerville R.P., Longpre J.-M., Jungers K.A., Engle J.M., Ross M.,			
RA	Evanko S., Wight T.N., Leduc R., Apte S.S.;			
RT	"Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS subfamily related to Caenorhabditis elegans GON-1.";			
RL	J. Biol. Chem. 278:9503-9513(2003).			
[3]				
RN	SEQUENCE OF 159-1935 FROM N.A. (ISOFORM 2).			
RP	TISSUE=Brain;			
RX	MEDLINE=20181126; PubMed=10718198;			
RA	Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro.";			
RL	DNA Res. 7:65-73(2000).			
CC	-!- FUNCTION: Cleaves the large aggregating proteoglycans, aggrecan and versican.			
CC	-!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 1838-Glu- -Ala-1839 site and versican at the 1428-Glu- -Ala-1429 site.			
CC	-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).			
CC	-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).			
CC	-!- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=3;			
CC	Name=1; Synonyms=ADAMTS-9B;			
CC	ISOId=Q9P2N4-3; Sequence=Displayed;			
CC	Name=2; Synonyms=long;			
CC	ISOId=Q9P2N4-1; Sequence=VSP_007548; VSP_007549;			
CC	Note=May result from the retention of an intron in the cDNA leading to a premature stop codon;			
CC	Name=3; Synonyms=Short;			

P58459 mus musculus  
Q07008 rattus norv  
P46531 homo sapien  
P35555 homo sapien  
O15072 homo sapien  
P11880 mus musculus  
Q04592 mus musculus  
Q8c919 mus musculus  
P07996 homo sapien  
Q61554 mus musculus  
P35448 xenopus lae  
P21783 xenopus lae

## ALIGNMENTS

Isoid-Q9P2N4-2; Sequence=VSP\_005499, VSP\_005500;  
-!- TISSUE SPECIFICITY: Highly expressed in all fetal tissues.  
Expressed in a number of adult tissues with highest expression in  
heart, placenta and skeletal muscle.  
-!- DOMAIN: The spacer domain and the TSP type-1 domains are important  
for a tight interaction with the extracellular matrix (by  
similarity).  
-!- DOMAIN: The ancillary domains, including the TSRs domain, are  
required for specific extracellular localization and for its  
versicanase and aggrecanase activities.  
-!- PM: The precursor is cleaved by a furin endopeptidase (by  
similarity).  
-!- SIMILARITY: Belongs to peptidase family M12B.  
-!- SIMILARITY: Contains 1 disintegrin-like domain.  
-!- SIMILARITY: Contains 1 GON domain.  
-!- SIMILARITY: Contains 15 TSP type-1 domains.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF261918; AAF89106.1; -  
EMBL; AF488803; AAO5765.1; -  
EMBL; AB037733; BAA92550.1; -  
HSSP; P15167; LATL.  
MEROPS; M12.021; -  
Genew; HGNC:13202; ADAMTS9.  
MIM; 605421; -  
GO; GO:0008237; F:metallopeptidase activity; TAS.  
GO; GO:0007275; P:development; TAS.  
GO; GO:0006516; P:glycoprotein catabolism; TAS.  
InterPro; IPR001762; Disintegrin.  
InterPro; IPR001818; Pept\_M10A\_M12B.  
InterPro; IPR006025; Pept\_M\_Zn\_BS.  
InterPro; IPR001590; Peptidase\_M12B.  
InterPro; IPR002870; Peptidase\_M12B\_N.  
InterPro; IPR000884; TSP1.  
Pfam; PF01562; Pfam\_M12B\_propep; 1.  
Pfam; PF01421; Reprolysin; 1.  
Pfam; PF00090; TSP1; 11.  
SMART; SMO0209; TSP1; 12.  
PROSITE; PS02015; ADAM\_MEROPS; 1.  
PROSITE; PS00546; CYSTEINE SWITCH; FALSE NEG.  
PROSITE; PS04227; DISINTEGRIN\_1; FALSE NEG.  
PROSITE; PS0214; DISINTEGRIN\_2; FALSE NEG.  
PROSITE; PS50092; TSP1; 14.  
PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
Repeat; Extracellular matrix; Alternative splicing.  
FT SIGNAL 1 18 POTENTIAL.  
FT PROPEP 19 287  
FT CHAIN 288 1935  
FT DOMAIN 293 499  
FT DOMAIN 509 587  
FT DOMAIN 588 643  
FT DOMAIN 644 752  
FT DOMAIN 753 877  
FT DOMAIN 878 936  
FT DOMAIN 939 997  
FT DOMAIN 998 1049  
FT DOMAIN 1052 1109  
FT DOMAIN 1110 1166  
FT DOMAIN 1182 1240  
FT DOMAIN 1241 1296  
FT DOMAIN 1328 1379  
FT DOMAIN 1382 1440  
FT DOMAIN 1441 1494  
FT DOMAIN 1497 1555  
FT DOMAIN 1556 1611  
TSP TYPE-1 1.  
TSP TYPE-1 2.  
TSP TYPE-1 3.  
TSP TYPE-1 4.  
TSP TYPE-1 5.  
TSP TYPE-1 6.  
TSP TYPE-1 7.  
TSP TYPE-1 8.  
TSP TYPE-1 9.  
TSP TYPE-1 10.  
TSP TYPE-1 11.  
TSP TYPE-1 12.  
TSP TYPE-1 13.

FT	DOMAIN	1612	1676	TSP TYPE-1 14.
FT	DOMAIN	1677	1734	TSP TYPE-1 15.
FT	DOMAIN	1735	1935	GON.
FT	DOMAIN	88	96	POLY-SER.
FT	SITE	223	223	CYS-STEINE SWITCH (POTENTIAL).
FT	METAL	434	434	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT SITE	435	435	BY SIMILARITY.
FT	METAL	438	438	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	444	444	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	112	112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	135	135	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	271	271	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	743	743	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	840	840	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1213	1213	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1267	1267	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1788	1788	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1806	1806	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	1064	1072	CLVTCGKGH -> VRWEGCYFP (in isoform 3).
FT	VARSPLIC	1073	1935	/FTid=VSP_005499.
FT	VARSPLIC	1624	1629	Missing (in isoform 3).
FT	VARSPLIC	1630	1935	/FTid=VSP_005500.
FT	VARSPLIC	1630	1935	CSVTG -> VPSWEL (in isoform 2).
FT	VARSPLIC	1630	1935	/FTid=VSP_007548.
FT	CONFLICT	46	46	Missing (in isoform 2).
FT	CONFLICT	96	96	/FTid=VSP_007549.
FT	CONFLICT	182	182	S -> G (IN REF. 1).
FT	CONFLICT	367	367	P -> S (IN REF. 1).
FT	CONFLICT	1117	1117	D -> G (IN REF. 2).
FT	CONFLICT	1117	1117	F -> L (IN REF. 1).
FT	CONFLICT	1117	1117	V -> G (IN REF. 3).
FT	SEQUENCE	1935	1935	AA; 216556 MW; FD3D51B88300A3C6 CRC64;
QY	201	WSS	201	WSS-----WSPCKTCTSGSLLPFRSRNRNKHMTAIGGKCEKPELLEKACIVEGELL--- 254
DB	881	WNSHGPWQACSKPCQ	881	WNSHGPWQACSKPCQ-----GERK-----KLVTRESQPLTWS 914
QY	255	QCPRYSWRSEWKEQVUSLLEQDQPHVHT	255	QCPRYSWRSEWKEQVUSLLEQDQPHVHT-----GPVGGGIQTRVYVCAQSVAA 308
DB	915	DQCDRLFPQGHITPCGTDC	915	DQCDRLFPQGHITPCGTDC-----DLRHWVARSSECSAQCGLYRTLDIYCAK----- 964
QY	309	AALRAKEVSRPVEKALCVGPAPLPOLCNI	309	AALRAKEVSRPVEKALCVGPAPLPOLCNI-----PCSTDCIVSSW-----SAWGLCIHENCHEPOCK 366
DB	965	YSRLDGKTEKVDGFC	965	YSRLDGKTEKVDGFC-----SSHPKPSN-----REKSGECNTGHWYSANWTEC----- 1009
QY	367	KGFRTRQRLVLMESTGPAGHCHPLVESVPC	367	KGFRTRQRLVLMESTGPAGHCHPLVESVPCEDPMCYRWLASEGICFPDHGKGLGHLRLK 426
DB	1010	-----	1010	-----SKSCDGTQRRR 1021
QY	427	AVCONDRGEDVSGSLCPVPPPPKSC	427	AVCONDRGEDVSGSLCPVPPPPKSC-----EIPCRMDCVLSEWT-----EWSSCSOSCSNKNSDG 483
DB	1022	AICVTRNDVLDKSKTHQEKVTIQRCEFC	1022	AICVTRNDVLDKSKTHQEKVTIQRCEFC-----PQKSGDWSECLVTC-----G 1069
QY	484	KQTRERTILALAGE	484	KQTRERTILALAGE---GGKPCPPSQALQEHRLNDSHSMQHLWETSHPGCPSEDTLWTA 540
DB	1070	KGKHKRQVWCQFGEDRLNDRMCDPTEKPT	1070	KGKHKRQVWCQFGEDRLNDRMCDPTEKPTSMQTCQPECAS---WOAGPWQCS----- 1120
QY	541	LNATIGNGEATCGVIGITRVRFC	541	LNATIGNGEATCGVIGITRVRFC-VKSHVGVQVMTKRCFDSRTPETVPCFLPCKKCG-- 597
DB	1121	-----	1121	-----VTGQGYQLRAVKCIIGTINVSVDNDNCNAATRTDTQDCSLP---SCHP 1167
QY	598	-----	598	-----IVTAFSEMTPCPRMCAQGNATVKQSRRIIIQEAANGQCEPDT 641
DB	1168	PPAAETRRSTYSAPRTQWRFGSNTWPCAT	1168	PPAAETRRSTYSAPRTQWRFGSNTWPCATCGK-----TMRVYVSCRDENG-----S 1215
QY	642	LYBEREC	642	LYBEREC-----EDVSLCPVTRWKPQKWSPCILLVPSVWQGITGSSACGKGLQTR 692
DB	1216	VADESACATLPRVAKRECSVTPCGQKALDWSS	1216	VADESACATLPRVAKRECSVTPCGQKALDWSS-----SVTCQGRATR 1261

## Query Match

8.1%; Score 664; DB 1; Length 1935;

Best Local Similarity 21.3%; Fred. No. 1.7e-36;

Matches 289; Conservative 123; Mismatches 42; Indels 526; Gaps 70;

```
693 AVSICSDNRSAMMECKLQTMGMPLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSR 752
b 1262 QWCVNYSVDHIDRSEC--DDYIETDQDCM-----SPC----- 1295
y 753 RQLTGKSKKKEKQSDLI--YPLVETELCPDEFISOPY-----GWSDCILPEG 801
b 1296 -----PORTPDSGLAQHFQNEYRPRASPSRTHVLGNQWRTGPGAC----- 1340
y 802 RREPHRLRVQADSKCEGSLGFRAVACDKNG-----RPVDFSCSSSSYIQEK 851
b 1341 -----SSTCAGSRRVVVVCODENGYTANDCVERIKDEQRAESG----- 1381
y 852 CVIPDFCKLSDWS--SWGSSSSCGIGVIRIRSKWLKEKPYNGRCPKLDLKNQVHEA 909
b 1382 ---PCP-----QWAGNWGECKLCCGGIRTR--LVVQKRSNGERFP--DLSCEILDK 1427
y 910 VPCYSECNOY-----SWVEHWSCKINNELRSRCCGGTQSRKIRCVNTADGEGGAVD 963
b 1428 PPDREOCNTHACPHDAWSTGPWSSC-----SVSCGRGHKQRYVYCMAX---DGSHLE 1477
y 964 SNLCNQDEIPEPTQSC--SLMCPNECVMSW--GLWSKCPQSCDPTMQRTRHLLRPSLN 1020
b 1478 SDYCKHLAKPHGRKCRGRCRCP-----KWKAGAWSQCSVSCGRGVQQRH---VGCCIG 1527
y 1021 SRTCAEDSQVQCLLINEN-----CFQFYNLTEWSTCQLSENAPCGQGVTRLLSCV 1072
b 1528 THKIARETECNPTYPRESRDCQGRCPPLYTWRAEWQEC-----TKTCGSGSRVKKVCV 1583
y 1073 -----CSDGK--PVSMDQCEQHNLEKPFQMSIPCLVCEVNVNQLSGWTA--WTEC 1118
b 1584 DDMQNEVHGARDVSKRPVDRSCSLQ-----PC--EYV-----WITGEWSEC 1624
y 1119 SQTCHGGRMSTRFIIMTQGEGRPCPELTQEXT--CP-----VTPCY-----S 1162
b 1625 SVTCGKGYKQRLVSCSEIYTGRN-----YEYGYQTINCPGTQPPSPVPCVLRCPVSAT 1680
y 1163 WVLGNWSACKLEGGDCGEGVQVRSLSVMHSGSISHAA----- 1200
b 1681 WVGWNGSCSV--SCGVGMSRVQCLTNEQPHLCHTDLKPBERKTCRVNVCPLPQ 1737
y 1201 -----GRVEDALCGE--MPFQDSILKQLCVPCPGDCHLTSEWSEWSTCLFCIDGR 1249
b 1738 NCKEVRKRLGASED--GEYFLMIRGKLLKIFCA-----GMHSDHPKEVVT--LVHGDSE 1787
y 1250 SP-ETVGRQSRSTFIIOFENQDSC-----PQQVLEI----- 1281
b 1788 NFSEVYGHRLHNPTECPYNGSRDDCCQKQVTAAGFSFQKIRIDLTSMQIITDLOFA 1847
y 1282 ----RP---CTGGKCYHYT-----WKASLWNN-----NERTVWCQRS-----D 1312
b 1848 RTSEGHVPVPFATAGDCYSAKCPQGRFSINLYGTGLSTESARWISQNGYAVSDIKKSPD 1907
y 1313 GUNVTGCGSPQAPPAIROCIACRKPFSYCTQGGVCG 1350
b 1908 GTRVWGC-----GGYCG 1920
```

## RESULT 2

```
AT20_MOUSE
ID AT20_MOUSE STANDARD; PRT; 1906 AA.
AC P59511;
JT 10-OCT-2003 (Rel. 42, Created)
JT 10-OCT-2003 (Rel. 42, Last sequence update)
JT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAMTS-20 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 20) (ADAM-
DE TS20).
EN ADAMTS20.
EN Mus musculus (Mouse).
XC Fukuyama; Metazoa; Chordata; Vertebrata; Euteleostomi;
XC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
DX NCBI_TaxID=10090;
[1]
```

```
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Fetal brain;
RX MEDLINE=22566039; PubMed=12562771;
RA Llamazares M., Cal S., Quesada V., Lopez-Otin C.;
RT "Identification and characterization of ADAMTS-20 defines a novel
RT subfamily of metalloproteinases-disintegrins with multiple
RT thrombospondin-1 repeats and a unique GON domain.";
RL J. Biol. Chem. 278:13382-13389(2003).
[2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND DISEASE.
RC STRAIN=DEA/2;
RX MEDLINE=22806432; PubMed=12925592;
RA Mac C., Foerzler D., Loftus S.K., Liu S., McPherson J.D.,
RA Jungers K.A., Apte S.S., Pavan W.J., Beier D.R.;
RT "A defect in a novel ADAMTS family member is the cause of the belted
RT white-spotting mutation.";
RL Development 130:4665-4672(2003).
CC -!- FUNCTION: May play a role in tissue-remodeling process occurring
CC in both normal and pathological conditions.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=ADAMTS20 B long isoform;
CC IsoId=P59511-1; Sequence=Displayed;
CC Name=2; Synonyms=ADAMTS20 A short isoform;
CC IsoId=P59511-2; Sequence=VSP_007606, VSP_007607;
CC -!- TISSUE SPECIFICITY: Expressed at low level in testis and brain.
CC -!- PFM: The precursor is cleaved by a furin endopeptidase (By
CC similarity).
CC -!- DISEASE: Defects in ADAMTS20 are the cause of the belted (bt)
CC phenotype. It is a pigmental defect which occurs as a result of a
CC defect in melanocyte development.
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.
CC -!- SIMILARITY: Contains 1 GON domain.
CC -!- SIMILARITY: Contains 15 TSP type-1 domains.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ512753; CAD54808.3; -
CC EMBL; AY189815; AAC74895.1; -
CC EMBL; AY189816; AAC74896.1; -
CC MGI; MGI:2660628; Adamts20.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR006025; Pept_M12B.
CC InterPro; IPR001590; Peptidase_M12B.
CC InterPro; IPR002870; Peptidase_M12B_N.
CC InterPro; IPR000884; TSP1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC Pfam; PF01421; Reprylisin; 1.
CC Pfam; PF00030; tsp_1; 11.
CC SMART; SMO0209; TSP1; 14.
CC PROSITE; PS00215; ADAM_MEPRO; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS0214; DISINTEGRIN_2; FALSE_NEG.
CC PROSITE; PS0092; TSP1; 13.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 26
FT PROPEP 27 249
FT CHAIN 250 1906
FT DOMAIN 250 464
FT DOMAIN 465 552
FT DOMAIN 553 608
FT TSP TYPE-1 1.
```

FT	DOMAIN	609	720	CYS-RICH.
FT	DOMAIN	721	842	SPACER.
FT	DOMAIN	843	901	TSP TYPE-1 2.
FT	DOMAIN	906	962	TSP TYPE-1 3.
FT	DOMAIN	962	1015	TSP TYPE-1 4.
FT	DOMAIN	1017	1074	TSP TYPE-1 5.
FT	DOMAIN	1075	1131	TSP TYPE-1 6.
FT	DOMAIN	1148	1202	TSP TYPE-1 7.
FT	DOMAIN	1203	1260	TSP TYPE-1 8.
FT	DOMAIN	1300	1351	TSP TYPE-1 9.
FT	DOMAIN	1354	1411	TSP TYPE-1 10.
FT	DOMAIN	1412	1465	TSP TYPE-1 11.
FT	DOMAIN	1468	1526	TSP TYPE-1 12.
FT	DOMAIN	1527	1584	TSP TYPE-1 13.
FT	DOMAIN	1585	1648	TSP TYPE-1 14.
FT	DOMAIN	1650	1706	TSP TYPE-1 15.
FT	DOMAIN	1707	1906	GON.
FT	METAL	399	399	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	400	400	BY SIMILARITY.
FT	ACT SITE	403	403	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	409	409	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	92	92	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	221	221	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	714	714	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	798	798	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	805	805	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1057	1057	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1562	1562	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1719	1719	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1759	1759	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1777	1777	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	1424	1425	CS -> VR (in isoform 2).
FT	VARSPLIC	1426	1906	/FTid=VSP_007606.
FT	CONFLICT	1211	1211	Missing (in isoform 2).
FT	CONFLICT	1262	1262	/FTid=VSP_007607.
FT	CONFLICT	1262	1262	D -> Y (IN REF. 2).
FT	CONFLICT	1262	1262	S -> L (IN REF. 2).
FT	SEQUENCE	1906 AA;	212041 MW;	0EB2568547B557EA CRC64;
Qy	Query Match	7.68;	Score 628.5;	DB 1; Length 1906;
Db	Best Local Similarity	21.11;	Pred. No. 3.9e-34;	
Qy	Matches 377;	Conservative 163;	Mismatches 558;	Indels 691; Gaps 96;
Qy	85	CDAPISCPGEE	----	YFSLKVG-----105
Db	377	CDPSRCSISENGLSAFTIAHGLGVNPHDSEKKEAGIKHQYHMAPTLNYHTS	----	-----105
Qy	106	--PWSKRLPHLKEI	----	-----NPSGRV-----134
Db	437	PWTWSACSQKHITFLDTGHGECILLDPKNGRTYDLSPLQFSGVYDGNRQCELMFGPGSQV	----	-----LDNDSNE 134
Qy	135	RVTFKHOSYKAHHHSKSWAIEIGYQTRQVSTSDGONAMLSLCLQDSFPLT-VQSC----	----	-----190
Db	497	CPYLKH-----	----	-----CARLWCTSAEGVH-----KGCRTQHMPLADGTCGPG 534
Qy	191	-----IMPKDCET-----SQWSSWSP-----CSKTCRSGSLIPG-----FRSRNRVKGMAIG 233	----	-----233
Db	535	MHCERGLCVTRDMETRPVDGSGWPGWPGYSSCSRCCGGGIKSTARLCLDRPEPRNGRCVVG 594	----	-----594
Qy	234	-----GGKECP-----ELLEKACIVEG-----	----	-----LLQOCPRYSW-----262
Db	595	RMKFRSNTDSCPKGRDPEKQCSDFDGHFINGLPNFWLPGKYSIGYAVKDRCKLY 654	----	-----654
Qy	263	-----RTSEWKECVQSL-----	----	-----LEQDDPHWHTGP-----VCGG-- 292
Db	655	CRVAGTTSFYQLKDRVADGTPCGTETNDICVQGLCRQAGCDHVLNSKAKEDKCGVCGDN 714	----	-----714
Qy	293	-----GIQTRVY-----CAOSVFAAA-----	----	-----309
Db	715	SSCOTLAGVFNSAHYGVNVVKIPAGATNIEILOHSYSGRPEDDNYLALSDTQGNFLLG 774	----	-----774
Qy	310	-----ALRAKEV-----	----	-----SRPEKALCVGPAPLPS--QLCN 337

Db	775	NFVWSMAKKEINIQAQVFEYSGSNISIRINSTDRLEAEVLQVLGVNLYNPDPVRYSFN	834
Qy	338	IPCSIDCVSSW---SANGLCIHENCHEPQCKGFRTRQRHVLMESTGPAGHCP-HLIVES	393
Db	835	IPIERSNLFSDPYGPDCT-KFOCGLHRRKACVRSKSHAVVSDHNCGLHMLPLFT	893
Qy	394	VPCEDPMCYRW-LASEGICFPDHGKGLGHRILKAVCON---DRGEDVSGSLCPVP----	445
Db	894	EKCNMDCELRWHIIGKSDC---SSQCGGYRTLDVHCWKYSVHKQAV-----PVGQYQC	945
Qy	446	-----PPPERKSCIEPCMDCVLSEW---TENSSCQSCSNKNSDKQTRSRITILAGEGG	499
Db	946	GDLKPPSRE---PCGSCVLTRWHYSEWSQCSRSR-----GGGDKTRESYCVNGFG-	994
Qy	500	KPCPQSALQOEHL-----CNDHSCMOLHMETSPWGPCSEDTLVTALNAT	544
Db	995	-----HRLAESRELPVLENCNEFFCPS---WATSEWSEC-----	1029
Qy	545	IGWGEATCGVGIQTRRVFCVSHVGVQVTKRCPDSTRPETVRPCFLPKCKDCIVTAFSE	604
Db	1030	-----PVTGCKMKQRQVWCQSE-DPMRDGFCNASTKPSLRPECL---RACASWHVGP	1080
Qy	605	WTPCPRMCAQGNATVQKSRVRIIIOEAANGQCECPDTLYBEREC-----EDVSLCP	655
Db	1081	WGSTATCGHG-----YQM---RAVKCISEIFGTMLDDRECPQASRPSDRQDCILAP	1129
Qy	656	-----VYRWKPKWSPCLLPVSVWQGITSSACGKGLQTRAVSC	696
Db	1130	CLAIPEVGATSLPAITPLGAAQWRHGSWTPC-----SVSCGRGSOARYVSC	1175
Qy	697	LSDDNRSAMMECLKQTMGMLLVQECTVPCREDCTTAM--SKFTPCSTNCEATKRRR	754
Db	1176	RDADEVADESNAHLPR--PAAVSLCFPCGE-----WQAGDWSPCSASGCHKTRR	1227
Qy	755	QLTGKSRK--KEKQDSDLYLVEETE-----LPCDEFISQPYGNWSDCILP-----	799
Db	1228	VLCVNVHQLVDESYPDGRPVTEQCSLAACP--PSYSRAPSSSEQSPHVPNRVPLTHK	1286
Qy	800	-----EGREPHRGLRVQ-----ADSKCEGELRFAVACSDKNR-----PVD	838
Db	1287	PGENDQQAQISIRGNQWRTGFWGACSSCAGGLOHRAVWCQDEBDRSATSCDSSSKPPE	1346
Qy	839	PSFCSSSGYIOEKCVIPCFPFCKLSWSSWGSSSCGIGVIRSRKMLKEPYNG----	894
Db	1347	SRHCSG-----PCPH---WNYGDWCECTQTCGGV--KSRFVICQFPNGQWTOE	1391
Qy	895	RPC--PKLDLKNQVH-EAVPCYSECNQYSVVEHWSCKINNELRLSRCGGTOGRKIRC	951
Db	1392	HSCELPKFPMMQCHLHACP-----EDVSWYRGPWKSC-----SASCGKGVYREVL	1439
Qy	952	VNTADGEGAYDSNLCNQBDEIIPETQSC-SLMCPNECVMSW--GLWSKCPQSCDPHTMQ	1008
Db	1440	IDQFORK--LEEKYCSHLHPRTHKACRSRCRPS-----WKANKWKECVTCGSGVQ	1490
Qy	1009	RTRHLLRPSLNSRTCAED-----SQVPCLLNENCQFQYNLTBWSTCOLSENAPCG	1061
Db	1491	REVCYRLR---GTGRVSEDMCDPSTRPOGQRCWRQCMRYQWTTGDWLDGSTS----	1543
Qy	1062	QGVTRFLLSCVSDGKPYSDMOCEQHNLKPKQMSIPLCEVCVNCOLSGWTATECSQT	1121
Db	1544	KKETVLLKVCNEQNVQANESLCPD--LTKP--LSI-----	1575
Qy	1122	CGHGRMSRTFIIPTQGEGRPCPTLTQKTCPTTPC-YSWVLGNWSACKLEGGDCGE	1180
Db	1576	-----KCRNPHCKYSVVTGDSQSC---AGNCGF	1601
Qy	1181	GVQIRSLSCVHSGSISHAAGRVEDALCGEPFQDSILKQLCSV-PCPGDCHLEW--SE	1237
Db	1602	TSPQKITCTYKIQSSKRTFHQLRPVYVYGECPVPSPOAYKCDLRSC---LHVATWKVK	1658
Qy	1238	WSTCELTCTIGRSFETVGRQGRSTFIIQSFEN---QDSCPOQVLETRPCGKCYHYTW	1294
Db	1659	NSKCSVTGIG---IMERRVACRT-----ENGWPSDLCLKEL---KPDQAQKCYA---	1702

```

1295 KASLMNNERTWVCR-----SDG-----VNVGTG-----GCSPQ--ARPAAIRQCIPAC 1336
1703 -----NDCKLLATCKELQVNTVNVTKGDYDLNVRGRILKIHCSGMQLENP---REYLPV 1754
1337 RKPFYSV-----CTGGV-----CGCEK-----GTEINKSGFLDYCKVKVPGS 1374
1755 KEDNPFSEIYGLRQNPVECPNGRRPRDCACENDYLPAGYVFSKVR-----VDLES 1808
1375 EDKQADV-----KNLSGKNRP-----VNSKIHDFKGSLSLQ-- 1405
1809 QIKTADLLFSQTLGSKAYFFATAGDCYSAARCPQGFSLNLAGTCMKISNTAK-WLAQGR 1867
1406 -----PLDPGRVKIYGVGGGFLMIFLFTSYLVCKKPKHQST 1448
1868 YASVVIHRSQDGTK---VYGRCGG-----FCGKCIPIHMAT 1899

RESULT 3
T20 HUMAN
C AT20 HUMAN STANDARD; PRT; 1911 AA.
C P59510;
T 10-OCT-2003 (Rel. 42, Created)
T 10-OCT-2003 (Rel. 42, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
DE ADAMTS-20 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 20) (ADAM-TS 20) (ADAM-
DE TS20).
IN ADAMTS20.
IS Homo sapiens (Human).
IS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
IS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
IS NCBI_TaxID=9606;
IS [1]
N SEQUENCE FROM N.A. (ISOFORM 1).
X MEDLINE=22513925; PubMed=12514189;
A Somerville R.P., Longpre J.-M., Jungers K.A., Engle J.M., Ross M.,
A Evanko S., Wright T.N., Leduc R., Apte S.S.;
T "Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS
T subfamily related to Canorhabditis elegans GON-1."
L J. Biol. Chem. 278:13382-13389(2003).
P [2]
N SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
P TISSUE=Liver;
X MEDLINE=22566039; PubMed=12562771;
X Llamazares M., Cal S., Quesada V., Lopez-Otin C.;
T "Identification and characterization of ADAMTS-20 defines a novel
T subfamily of metalloproteinases-disintegrins with multiple
T thrombospondin-1 repeats and a unique GON domain."
L J. Biol. Chem. 278:13382-13389(2003).
C -!- FUNCTION: May play a role in tissue-remodeling process occurring
C in both normal and pathological conditions.
C -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
C -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
C matrix (By similarity).
C -!- ALTERNATIVE PRODUCTS:
C Event-Alternative splicing; Named isoforms=2;
C Name=1;
C IsoId=P59510-1; Sequence=Displayed;
C Name=2;
C IsoId=P59510-2; Sequence=VSP_007106, VSP_007107, VSP_007108;
C -!- TISSUE SPECIFICITY: Very sparingly expressed, although is detected
C at low levels in testis, prostate, ovary, heart, placenta, lung
C and pancreas. Overexpressed in several brain, colon and breast
C carcinomas.
C -!- PTM: The precursor is cleaved by a furin endopeptidase (By
C similarity).
C -!- SIMILARITY: Belongs to peptidase family M12B.
C -!- SIMILARITY: Contains 1 disintegrin-like domain.
C -!- SIMILARITY: Contains 1 GON domain.
C -!- SIMILARITY: Contains 15 TSP type-1 domains.
C This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF488804; AAC15766.1; -.
CC EMBL; AJ515153; CAD56159.3; -.
CC EMBL; AJ515154; CAD56160.2; -.
CC Genew; HGNC:17178; ADAMTS20.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR006025; Pept_M_2n_BS.
CC InterPro; IPR001590; Peptidase_M12B.
CC InterPro; IPR002870; Peptidase_M12B_N.
CC InterPro; IPR000884; TSPI.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC Pfam; PF01421; Reptolysin; 1.
CC Pfam; PF00030; TSP_1; 11.
CC SMART; SM00209; TSPI; 12.
CC PROSITE; PS02115; ADAM_MEPPO; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS02114; DISINTEGRIN_2; FALSE_NEG.
CC PROSITE; PS00092; TSPI; 12.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 21
FT PROPEP 22 253
FT CHAIN 254 1911
FT DOMAIN 254 467
FT DOMAIN 468 555
FT DOMAIN 556 611
FT DOMAIN 612 723
FT DOMAIN 724 846
FT DOMAIN 847 905
FT DOMAIN 906 962
FT DOMAIN 967 1024
FT DOMAIN 1025 1074
FT DOMAIN 1077 1136
FT DOMAIN 1153 1207
FT DOMAIN 1208 1265
FT DOMAIN 1305 1357
FT DOMAIN 1359 1417
FT DOMAIN 1418 1476
FT DOMAIN 1477 1532
FT DOMAIN 1536 1589
FT DOMAIN 1590 1653
FT DOMAIN 1655 1711
FT DOMAIN 1712 1911
FT METAL 403 404
FT ACT_SITE 404 404
FT METAL 407 407
FT METAL 413 413
FT CARBOHYD 92 92
FT CARBOHYD 191 191
FT CARBOHYD 445 445
FT CARBOHYD 702 702
FT CARBOHYD 717 717
FT CARBOHYD 728 728
FT CARBOHYD 809 809
FT CARBOHYD 870 870
FT CARBOHYD 1062 1062
FT CARBOHYD 1457 1457
FT CARBOHYD 1543 1543
FT CARBOHYD 1573 1573
FT CARBOHYD 1764 1764
FT CARBOHYD 1782 1782
FT CARBOHYD 1853 1853
FT VARSPLIC 683 693
FT FTID=VSP_007106.
CSASCCKGRKYREVFCDQFKLEDTNCSQVKPPTHKAC

```

```
FT RSVPCSKWANSMECSVTGCGVQQRDVYCRILK -> EDL
FT KXLLPQRTTILWELMKNIIFCHGSHMYLINVVTDLHLYP
FT RHCDPTEIETVFLSLWSLQFTWGLDKYXKNSL (in
FT isoform 2).
FT /FTid=VSP_007107.
FT Missing (in isoform 2).
FT /FTid=VSP_007108.
FT T -> V (IN REF. 2).
FT E -> Y (IN REF. 2).
FT T -> M (IN REF. 2).
FT D -> E (IN REF. 2).
FT I -> V (IN REF. 2).
FT MISSING (IN REF. 2).
FT LILQ -> ILIE (IN REF. 2).
FT GMLLAK -> WHVIG (IN REF. 2).
FT S -> Q (IN REF. 2).
FT T -> R (IN REF. 2).
FT T -> A (IN REF. 2; CAD56159).
FT T -> O (IN REF. 2; CAD56159).
FT 1882
SQ SEQUENCE 1911 AA; 214656 MW; CF592E220D32B250 CRC64;

Query Match
Best Local Similarity 21.6%; Pred. No. 1.1e-33;
Matches 306; Conservative 161; Mismatches 482; Indels 468; Gaps 73;

40 VVSFLPW--SNCKGCGKGLQHRTRAVIAP-PLFGGLQC-PNLTSRACDAPISCLPLG 94
555 VNGWGPWEPYSSCSRTCCGIEIATRCNPEPRNGYVCGVRMKFRSCNTD-SCPKG 613
95 BEEVTFSLKGVFWSKCLPLHLKINPGRVTLDFNSDSNERVTFKHSYKAHHHSKSWAI 154
614 TQDF-----REKQCSDFNG-----KHLDISGIPSNVRLP 643
155 E--IGYQTR-----QVSCTR-----SDQVAMLSLCLQDSFPLTVQSCIMPKDCE 197
644 RYSGIGTKDRCKLVCQVAGTNYFYLLKDMVEDG-----TPCGTETHDICVQGCMAAGCD 698
198 -----TSQSSWSPSCSKTCRSGSLPGRFSR-----NYGMAIGGK 236
699 HVLNSSAKIDKCGVCGGDSCKTITGVFNESHVYVWVKIPAGATNVDIRQVYSYSGQP 758
237 ECPLELEKACIVGELLQOCPRYSWRTSEWKEQVSLLEQDPHWHVTVGPVCGGIQT 296
759 DSYLAJSDA--EGNFL-----FNGNFLSTSKKEINVQ-----T 792
297 REVYCAQSVPAALRAKEVSRP-----VEKALCVGPAPLPS--QLCNIPGCTDCIVSSWS 350
793 RTV-IEYSGSNNAVERINSTNRQEKELLQVLCVGNLYNPVHYXFNIPLEERSDMFTWD 851
351 AWG-----LCIHENCHEPQCKGFRTRQRHVLVLMESTGPAGHCPHL----- 390
852 PYGPWEGCTKVCQGLQRNITCIHKS DHSVVS DK-----ECDHLPLP 893
391 ---VESVPECDPMCYRLASGICFPDHPGKGLGHRILKAVQN---DRGE--DVSGSLC 442
894 SFVTSQNTDELGRMLLAKSEK-----SSQCGQYRTLDHCKYSHIEGQTVQVDHYC 950
443 ---PVPPPERKSCIEPCRMDCVLEW---TEWSSCSQCSNKNKSGKTRRTILALAGE 498
951 GDQLKPPTQEL-----CHGNCVFRWHYSEWSQCSRSK-----GGGRSRESYCMNFG 999
499 GKPCPSQALQEHRL-----CNDSHCQMLHWETSPWGPCSDETLVLTALNA 543
1000 -----HRLADNEQELSRVTRENCNEFSKPS--WAAENSECL----- 1035
544 TIGMNGEATCGVGIQTRVFCVSHGVGMVKTRCPDSTRPETVFPCLPKCKDCIVTAFS 603
1036 -----VTCGKGTQRQVWC--QLNVDLSDGFCNSSTKPSLSPCEL---HTCASQVG 1084
604 BWTPCRMCOAGNATVKSRVRIIIQEAANGQGCPTLYEERECEDVS----- 652
1085 PMGPGTTCGCHG-----YQMRDVKCVN---ELASAVLEDTECHASRPSDRSQSVLT 1133
```

```
QY 653 -----LCP-----VYRWKPKQKSPCILLVPBSWVGITGSSBACGKJQTRAVS 695
DB 1134 PCSFTSKLETALLPVLIKKQAQWRHGSWTFC-----SYSCGRGTQARTVS 1179
QY 696 CJSIDNRSRAEMMECLKQTNMGVELLVQECTVPCRDECTETAW--SKFTPCSTNCEATKRR 753
DB 1180 CRDALDRADIESYCHLPR--PAEINWCFTCGE-----WQAGDWSPCSASCGHKTR 1231
QY 754 RLQTKSRK--KEKQDSLDLYPLVETE-----LCP--CDEFISQPY----- 790
DB 1232 QVLCMNYHQPIDENYCDPEVRPLMBEQECLAAACPAHSHFPSPVPQPSYLLTNLPLTK 1291
QY 791 --GNMSDCILPGRREPHRGLRVQADSKCEGELFRFRAVACDKNGRPVDPSPFCSSSG-- 846
DB 1292 LEDNENQVHVSVRGNQWRTGPFWSGSCSSCGGQHTAVVCCDENGQ--SASYCDAASKP 1349
QY 847 -YIQEKVIPCFDFCKLSDWS--SWGSSSSCGIGVIRIRSKWLKEKPYNGRCPCKLDLK 903
DB 1350 PELQCGGPGPCP-----QWYGNWGECSQTCGGIKSR--LVICQFPNGQILE--DHN 1398
QY 904 NOVHEAVPCYSECNOY-----SWVEHWSSCKNNNELRSRLRCGGTOSRKIRCUNTAGD 957
DB 1399 CEIVNKPSPSVIQCHMHACPADYSWHQEPWTSK-----SASCGKREYREVFCIDQFOR 1451
QY 958 EGGAVDSNLCNDEIPEPTQSC--SLMCPNECVMSEW--GLMSKCPQSCDPHTMORTRHL 1014
DB 1452 K--LEDTNCQVQKPPTHKACRSVRCPS-----WKANSNNECSVTGCGVQQRDVYCR 1502
QY 1015 LR--PSLNSRTCAEDSQVQPC-----LLNENCFQOYNLTENSTQLSNAPCGQGVTRL 1068
DB 1503 LKGVQVVEEMC--DQSTRPCSRRCWQDCVQH-----KGMERGR 1541
QY 1069 LSCVCSGDKGPMVMDCEQHNLKPKORMSIPCLVECVN----- 1106
DB 1542 LNCSTG-----CERK--DSHQM-----ECTDNQIRQVNEIVNYSSTISLTKNC 1584
QY 1107 ----COLSGWTA-WTECSQTCGHGRMSRTRFIIMPTQGBRCPCTEITQKTCVTP-- 1159
DB 1585 RNPPCNVIWVADSQCANNCGFSYRQITCTEIPSTKHKHLRLPIVYQECVPWPS 1644
QY 1160 -----CY-----SWVLGNWSACKLEGDCGEGVQIRSLSCMWHSGSI SHAAGRVEDALC 1208
DB 1645 QYQICNSCLHATWKVGNWSKCSV--TCGIGIMRWQVKCITKHGLSS----- 1690
QY 1209 GEMPFQDSILKCLSPVPCFG---DCHLTSEWSEWSTCE 1242
DB 1691 -----DLCLNHL-----KPGAQKCYANDCKSFCTCK 1717

RESULT 4
FSPO_RAT
ID FSPO_RAT STANDARD; PRT; 807 AA.
AC P35446;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE F-spondin precursor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic floor plate;
RX MEDLINE=92208952; PubMed=1555244;
RA Kiar A., Baldassare M., Jessell T.M.;
RT "F-spondin: a gene expressed at high levels in the floor plate
RT encodes a secreted protein that promotes neural cell adhesion and
RT neurite extension."
RL Cell 69:95-110(1992).
CC -!- FUNCTION: Promotes the attachment of spinal cord and sensory
CC neuron cells and the outgrowth of neurites in vitro. May
CC contribute to the growth and guidance of axons in both the spinal
```

```
CC cord and the PNS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.
CC -!- SIMILARITY: Contains 6 TSP type-1 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88469; AAA41174.1; -.
CC PIR; A38152; A38152.
CC InterPro; IPR002861; Reeler.
CC InterPro; IPR000884; TSP1.
CC Pfam; PF02014; Reeler; 1.
CC Pfam; PF00090; TSP1; 6.
CC SMART; SM00209; TSP1; 6.
CC PROSITE; PS50092; TSP1; 6.
CC Glycoprotein; Signal; Repeat; Cell adhesion.
FT SIGNAL 1 28
FT CHAIN 29 807
FT DOMAIN 442 495 TSP TYPE-1 1.
FT DOMAIN 501 555 TSP TYPE-1 2.
FT DOMAIN 558 611 TSP TYPE-1 3.
FT DOMAIN 614 666 TSP TYPE-1 4.
FT DOMAIN 668 721 TSP TYPE-1 5.
FT DOMAIN 754 806 TSP TYPE-1 6.
FT CARBOHYD 214 214 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 681 681 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 807 AA; 309525F9EAE9A9A CRC64;

Query Match
Best Local Similarity 5.2%; Score 428.5; DB 1; Length 807;
Matches 133; Conservative 64; Mismatches 158; Indels 107; Gaps 27;

2Y 849 QEKCVIPCPDCKLSDWSSGSCSS-CGIGVIRSKWLKEKYNGRCPKLDLKNQVH 907
3b 432 EEKDDPTETCIYSNWPSACSSSTCEKGMQRMLKQK-LDLSVCP-----DTQ 484
2Y 908 EAVPCY-----SECNOYSVWVHEWSSCKINNELRLSCGGGTQSRKIRCVNTADGE 958
3b 485 DFQPCMGPGCDEGSGCTWSEIT--WSPC-----SVSCGMGWSRE-RYVKQFPED 534
2Y 959 GGAVDNLNQDIP-PEIOSCSL---MCPNECMSEWGLWSKCPQSCDPHTMQRTREL 1014
3b 535 G-----SVC---MLPTETEKTWNBECSPPSCLVTEWGEWDDCSATCGMKKRRMYK 586
2Y 1015 LRPLNSRCAEDSQVQPCLLNENCFQVNL---TEWSTCOLSENAPCGQVGT--RL 1069
3b 587 MSPADGSMCKAETSAQKMWPE-CHTIPCLLSFWSEWSDCVT---CCKGMRTRQML 641
2Y 1070 SCVCSGDKPMSQCEGHNLEKFORMIPCLVCEVNVCLSGWTAWTECSQTGCHGRMS 1129
3b 642 KSLAELG-----DCNE-DLEQAEKMLP---ECPIDCELSEWSQSECNCKGK-GHMI 690
2Y 1130 RTFIEMPTGEGRCPCTELTOEKTCPVTFCYSWVLGNWSACKLEGDCGEGVQIRLSLC 1189
3b 691 RTTIQMEPFGGAPCP-ETVQKCKRARKC-----LRSPSIQKLWRARE----- 736
2Y 1190 MVHSGSISHAAGRDVALCCEMPFQDSILKQLCGVPCPGDCHLTSEWSTCELTICIDG- 1248
3b 737 -----SRSEQLREBSDEQ-----PPG-CRMRPTAWSECTKCGGGL 774
2Y 1249 -RSFETVGRSRSRTFTIIQSFENQDSCPQVLETRPCTGGKC 1289
3b 775 QERYMTVTKRFFKFSQF-----TSCKDK-KSIRACNVHPC 807

RESULT 5
FSPO_XENLA
```

```
FSPO_XENLA STANDARD; PRT; 803 AA.
P35447;
01-JUN-1994 (Rel. 29, Created)
01-JUN-1994 (Rel. 29, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE F-spondin precursor.
OS Xenopus laevis [African clawed frog].
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=93376785; PubMed=8367492;
RA Ruiz i Altaba A., Cox C., Jessell T.M., Klar A.;
RT "Ectopic neural expression of a floor plate marker in frog embryos
injected with the midline transcription factor Pintallavis.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8288-8272(1993).
CC -!- FUNCTION: Promotes the attachment of spinal cord and sensory
CC neuron cells and the outgrowth of neurites in vitro. May
CC contribute to the growth and guidance of axons in both the spinal
CC cord and the PNS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in the floor plate.
CC -!- SIMILARITY: Contains 6 TSP type-1 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L09123; AAA19105.1; -.
CC PIR; A47723; A47723.
CC InterPro; IPR002861; Reeler.
CC InterPro; IPR000884; TSP1.
CC Pfam; PF02014; Reeler; 1.
CC Pfam; PF00090; TSP1; 6.
CC SMART; SM00209; TSP1; 6.
CC PROSITE; PS50092; TSP1; 6.
CC Glycoprotein; Signal; Repeat; Cell adhesion.
FT SIGNAL 1 23
FT CHAIN 24 803
FT DOMAIN 436 489 TSP TYPE-1 1.
FT DOMAIN 495 549 TSP TYPE-1 2.
FT DOMAIN 552 605 TSP TYPE-1 3.
FT DOMAIN 608 662 TSP TYPE-1 4.
FT DOMAIN 664 717 TSP TYPE-1 5.
FT DOMAIN 750 802 TSP TYPE-1 6.
FT CARBOHYD 210 210 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 677 677 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 803 AA; 90702 MW; D3A54E32954BAED9 CRC64;

Query Match
Best Local Similarity 4.8%; Score 393.5; DB 1; Length 803;
Matches 132; Conservative 71; Mismatches 170; Indels 185; Gaps 25;

QY 760 SRXKEKQ-----DSDLYPLVETELCPDEFISQYGNW-----SDCILPEGRREPHR 807
3b 403 ARKEGQCNIIPDNVDDIVADLVTEKEDDTPETCIYSNWPSACSSATCDKGRVQR 462
QY 808 GLRVQADSKEGEGRLFRVAVACDKNGRPVDPSPCSSSGVIGKEKVIP-CPFD---CKL 862
3b 463 MLKQLD-----LSVPCPTQD-----FECMGPGGSDDEASTCM 498
QY 863 SDWSWGSCSSCGIGVIRSKWLKEXYNGRCPKLDLKNQVHEAVPCYSECNQYSW 922
3b 499 SEWITWSPCSASCWGIEVREYVYKQFPEDG----- 529
QY 923 VEHWSCKINNELSLRCGGGTQSRKIRCVNTADGEGAVDSNLCNDEIPEPTQSCSL- 981
```





46 PMSNCKGCKGLQHRTRAVIAP-PLFGGLQCP-----NLTESBAC--DAP-----ISC- 91  
550 PMSHCRTCGAGVQSAERLCNNPEPAGFGKYCTGERKVRLCNVHPCRSEAPTFQMQCS 609  
92 -----PLGEEVTFSLKVGPSKCRPLPHLKEIN-----PSGRV- 125  
610 EBDTVPKNELVHFFIFNPAHPCEL-YCRPIDGOFSEKXLDVADIDGTFPFEGGNSRNV 668  
126 -----LDFNSDSNE-----RVTFKHQSYKAHHSHKSMAIEIGY 158  
669 INGIKRWGCDYEDSNATEDRCGVCLGSSCQTVRKMFKEKESGY-----VDIGL 721  
159 -----OTRQVSCSTRSDGQVAMLSLQCP----- 181  
722 IPKGARDIRVMEIEGAGNFAIRSEDEPKYIYNGGFIQIOMNGYKLAGTVFOYDRKGBLE 781  
182 -----SFPLTVOSC:MPKDCETS-----QWSSWSPCKSTCRS 213  
782 KLMAOTPTNESVWIIQLFQVTFPIKYEYTIQKDGLDNDVEQMYFWQYGHWTCECVTGT 841  
214 GSLLPGRSRNVKH-----MAIGGKCEPPELLEKEACIVEGELLQOQPR 259  
842 GIR-----ROTAHCKIKRGMVATFCDPETQPNRGOKKCHE-----KACPP 883  
260 YSWRTSEWKECVSLLEQDDPHVHTGFCVGGGQTREVCYCAQSVPAALRAKEVSEP 319  
884 -RWAGENEACSAIC-----GPH-----GEKRTVLCTQM-----VSDQALPP 922  
320 VEKALCVGAPLPSQCN-IPFSTDCIVSSAMGLCIHENCHEPQGGKGFRTQRHVL 377  
923 TDCQHLKPKTLLS--CNRDILCPSDWTGNNS-----ECVSQCG-GGVRI- 966  
378 MESTGPAGCHPLVESVPC-----EDPMC--YRWLASGICFPDGHGKGLHRLIKAVQND 432  
967 -----SVTCAKNHDEPCDVTXKPNRGCALC--GLQCPSPSRVLKP--N 1005  
433 RGEDVSG-----SILCPVPPPPER-KSEICPRMDCVLSWTESSQSCSNKNSD--GKQ 485  
1006 KGTISNGKPNPILKVPVPTSPRMLTTTGESMSTSTPAISSPPTASKEGLDGKQ 1065  
486 TR-----SRTILAGEGKCPGPSAL-----OEHRLNDHSCMLHWTSPWGPCS 533  
1066 WQDSSTQPELSRYLISTGSTSQPILTSQSLIQPSENV-----SSSDTGTS 1114  
534 EDTLVATNALIWNGEATCGVGIOTRR-----VFCVSHVCQ----- 571  
1115 EGLVAT-----TTSGSLSSRNPIWTPVTPVNTLTKPEMEIHSGEEREQ 1164  
572 -----VMTK-RCPDSTRP--ETVRPCLPCKKDCIVTAFSEWTPC----- 608  
1165 PEDKDBSNPVIWTKIRVGNDAVVESTEMPLAPLTPD--LSRESWPPFSTVMEGLLPS 1222  
609 --PRMCQAGNATVKSRYLIIICEAN-----GGQECPTLYEERECEDEVSLCPVYRWKP 661  
1223 QRTTSETGTPE-----GMVTEKPANTLPLIGGHHQ----- 1256  
662 QKWSPCILVPESWQIGTSSSEACGKGLQTRAVSCISDDNRSAMWECILKQTMGMLLVQ 721  
1257 -----EPGSKTANRNLKLPNNNNQTKSEPVLTEDATSLITE 1295  
722 ECTVPCREDCFTT-----ANSKFTPTSTNCEATKSRRLQTLGSKRKEK--Q 767  
1296 GFLNANASYKQLTNGHGSARHIVGNWSE-----CSTTCG-----LGAYWKRVCFTTQM 1343  
768 DSDLY-----PLVETELCPCEFTISQPYGNNSDCILPEGREPHRGLRVQADSCEGEG 821  
1344 DSDCAAIQRPDPKACHLAPC--AGWKVGNWSK-----SNCSSG 1382  
822 LRPRVACSD-KNGRPVDFSFSSSYIIEKCVIPC-PFDCKLSDWSSWGSCSSCGIGV 879  
1383 FKIREIQCVDSRDLRNLRFPHQFLAGIPPLPLSMSCNPEPCAWQVWPSQCSRSRGGV 1442  
880 RIRS-----KWLKEKPYNGRCPCKLDLKNQVHEAVPCYSECNQY---SWVVEHWS 928

Db 1443 QERGVFCGGLCDWTK-----RPTSTM-----SCNEHLCCHWATGNWDL 1481  
Qy 929 KKNELRSRCCGGTQSRKIRCVNTADGEGGAVDSNLQNDLIPETQSCS----- 980  
Db 1482 C-----STSCGGGFKRIVQCVSEGNKTEQDQCLDKHPRPEFFKCNQOACKSA 1534  
Qy 981 -LMCPN-----ECVMSWGLMSKCPQSCDPTMORRTHLL 1015  
Db 1535 DLLCTDKLSASFQTLKAMKCSVPTVRAECCFS-----CFQTHITHTQRRQRLL 1587  
Qy 1016 RPS 1018  
Db 1588 QKS 1590  
RESULT 7  
SM5B MOUSE  
ID SM5B\_MOUSE STANDARD; PRT; 1093 AA.  
AC Q60519;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Semaphorin 5B precursor (Semaphorin G) (Sema G).  
GN SEMA5B OR SEMAG OR SEMG.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STEAIN-NMRI;  
RX MEDLINE=96414430; PubMed=8817451;  
RA Adams R.H., Betz H., Puschel A.W.;  
RT "A novel class of murine semaphorins with homology to thrombospondin  
is differentially expressed during early embryogenesis.";  
RL Mech. Dev. 57:33-45(1996).  
CC -!- FUNCTION: May act as positive axonal guidance cues.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: In adult, only detected in brain.  
CC -!- DEVELOPMENTAL STAGE: Differentially expressed in embryonic and  
adult tissues. Its abundance decreases from E10 to birth.  
CC -!- SIMILARITY: Belongs to the semaphorin family.  
CC -!- SIMILARITY: Contains 1 Sema domain.  
CC -!- SIMILARITY: Contains 7 TSP type-1 domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X97818; CAA66398.1; -;  
DR MGD; MGI:107555; Sema5b.  
DR InterPro; IPR003659; Plexin-like.  
DR InterPro; IPR002165; plexin\_repeat.  
DR InterPro; IPR001627; Sema.  
DR InterPro; IPR000884; TSPI.  
DR InterPro; IPR008085; TSP\_1.  
DR Pfam; PF01437; PSI; 1.  
DR Pfam; PF01403; Sema; 1.  
DR Pfam; PF00090; tsp\_1; 5.  
DR PRINTS; PR01705; TSP1REPEAT.  
DR SMART; SM00423; PSI; 1.  
DR SMART; SM00630; Sema; 1.  
DR SMART; SM00209; TSPI; 4.  
DR PROSITE; PS00092; TSPI; 5.  
KW Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;  
KW Developmental protein; Glycoprotein.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1093 SEMAPHORIN 5B.

FT DOMAIN 20 978 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 979 999 POTENTIAL.  
FT DOMAIN 1000 1093 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 336 518 SEMA.  
FT DOMAIN 551 605 TSP TYPE-1 1.  
FT DOMAIN 606 662 TSP TYPE-1 2.  
FT DOMAIN 664 713 TSP TYPE-1 3.  
FT DOMAIN 721 776 TSP TYPE-1 4.  
FT DOMAIN 795 850 TSP TYPE-1 5.  
FT DOMAIN 852 907 TSP TYPE-1 6.  
FT DOMAIN 908 952 TSP TYPE-1 7.  
FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 539 539 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1093 AA; 120326 MW; 29B5C9B1E8108717 CRC64;  
Query Match 4.4%; Score 358.5; DB 1; Length 1093;  
Best Local Similarity 22.4%; Pred. No. 2.3e-16;  
Matches 180; Conservative 66; Mismatches 248; Indels 309; Gaps 41;  
QY 579 DSTRPETVRPFLPCKKDCIVTAFSEWTPCPRMCGNATVKQRYRIIOEAAAGGQEC 638  
DB 395 EAVQVTEPC-----VTQDSVRFSLVDLV---QA 423  
QY 639 PDLTYEE-----REC--EDVSLCPVYRWKPKQNSCIIPLVPSVW 675  
DB 424 KDTLVHLYIGTESITLKALSTASRSGCYLBELHVLPPGRLEPLRSILHSARALF 483  
QY 676 QGITG-----SSEAC-----GKGLQTRAVSCISPDNRSAEWELCKOTNMPLLVQECT 724  
DB 484 VLSRVRIRIPLERSAVHSQACLGARDPFGWDGK--QLCTLEDSSNWSLWQNIIT 541  
QY 725 V-PCR---EDCTFTAWSKFTPOSTNCEATKRRRLTGKSRKSKQSDLIPLVETELC 780  
DB 542 TCPVRNVRDGGFGPSPWKPC-----EHLDDGNSGSLCRARS-----580  
QY 781 PDDEFISQPYGNWSDCIIPEGREPHRGLRVQADSKCEGELRPVAVACSKNGRPVDPDS 840  
DB 581 -CDS-----PRPR-----CG-----GLECL---GFSIHA 601  
QY 841 FCSGSGYIQEKVPCPFCDKLSWSSWGSCSSCGIGVIRIRSKWLKE-KPYNGGRPC-P 898  
DB 602 NCSRNG-----AWTASSWAQCSSTSGIGFQVRQSRCSNPAPRHGGRICVG 647  
QY 899 KLDLKNQVHEAVPCVSECNQVSVVHVHSSKINNELSLRCGGTQGRKIRCVNTADGE 958  
DB 648 KREERFNCENTPCVPPIFWASW--GSWSKCSNN-----CGGVSQSRSSCENGNSCP 698  
QY 959 GGAVDSNLCNDEIPEPTQSCSLMCPNECVMSWGLW-----SKCPQSC---1002  
DB 699 GCGVEFKTN-----PEA-----CPEVRNTPWTPMLPVAVTQGGARQGRFRTCRAP 747  
QY 1003 --DPHTMQ--RRTHLLRPSLNSRTCAEDSQVQCLLNENCFOFQYN-----ITEWSTC 1052  
DB 748 LPDPHGLQFGKERTETRTCPADGTGACDITDALVEDLLRSGSTSPHTLNGWATWGPWSSC 807  
QY 1053 QLSENAPCGQVTRTLLSCVSDGDFVSMQDCEQHENLEKPKQMSIPCLVE-----1102  
DB 808 -----SRDELGLRVRKRTCT-----NP-----EPRNGGLPUCVGDAAEYQDCNPQ 847  
QY 1103 -CVVNCQLSGMTAWTECSQTQGHGRMSTRPIIIMPTQGEGRP-CPTBLTQKTCPTVPC 1160  
DB 848 ACPVRGAWSCMTAWSCGASGCG-GGHYQRTSRCTSPAPSPGEDICLGLHTEALCSTQAC 906

QY 1161 -YSWVLGNWSACKLEGGDCGE-GVQIRSLSCWVHSGSISHAAGRVEDALCGEMPPQDSIL 1218  
DB 907 PEGMSL--WS-----EWGVCETDGAQSRSRSC-----EELL-----935  
QY 1219 KOLCSVPCPGDCHLTSEWSEWSTCELTICDGRSFETVGRQSRRTFIIOSEFNQDSCP---1275  
DB 936 -----PGFGAC-----VGNSSQSR-----PCPYSE 955  
QY 1276 -----QQVLETRPCTGGKCYH 1291  
DB 956 IPVILPASSVEITSCGGFNLIH 978  
RESULT 8  
BAR3\_CHITE STANDARD; PRT; 1700 AA.  
ID BAR3\_CHITE AC Q03376;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Balbiani ring protein 3 precursor.  
OS BR3.  
OS Chironomus tentans (Midge).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;  
OC Chironomidae; Chironominae; Chironomus.  
OC NCBI\_TaxID=7153;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Salivary gland;  
RA MEDLINE=90172404; PubMed=1689777;  
RA Paulason G., Lendahl U., Galli J., Ericsson C., Wieslander L.;  
RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged  
repetitive structure split by many introns."  
RL J. Mol. Biol. 211:331-349 (1990).  
CC -!- FUNCTION: Used by the larvae to construct a supramolecular  
role as a transport protein that binds to other proteins  
intracellularly and in the gland lumen in order to prevent these  
from forming water-insoluble fibers too early.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Salivary gland.  
CC -!- DOMAIN: Has 82 approximate repeats of Cys-x-Cys-x-cys.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; X52263; CAA36506.1; --  
PIR; S08167; S08167.  
HSP; P15358; 1SKZ.  
DR InterPro; IPR004153; CXXC-repeat.  
DR Pfam; PF03128; CXXC; 71.  
DR Repeat; Signal.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 1700 BALBIANI RING PROTEIN 3.  
SQ SEQUENCE 1700 AA; 186145 MW; 34202B28521B0815 CRC64;  
Query Match 4.3%; Score 357.5; DB 1; Length 1700;  
Best Local Similarity 19.6%; Pred. No. 4.5e-16;  
Matches 303; Conservative 146; Mismatches 595; Indels 499; Gaps 89;  
QY 28 EQACLIPCR--DCVVSEPLPWSNCKGCKLQHRTRAVIAPPLFGGLQCPLNLTESRAC 85  
DB 127 EKSCACVCFNAKCTAPQVNNKDTCCCGCFVNNQBPADGCTKPLIWDKVC-----RC 179  
QY 86 DAPISCPGLGEEYTFSLKVGWPS-----KCRPLHLKEINP 120

Db 180 ECPLKDKGNR-----DWSSSCSCECKGDKGKQSGKIWKKNCRICP---TAPF 228  
Qy 121 SGRVLDFNSDSNE-----RVTFKHQSYKAHHHSKW---AIEIGYQTRVOSTRSDGON 172  
Db 229 AGGCSAPLKWDKSCACAPAKMEKKEKCVESKIWNPNPTCEG--CAQLNCPDNKKAN 286  
Qy 173 AMLSLCLQDSFPLTVQSCINPKDCETISQNSWSWPCS-----KTCRSGSLLPGRFSR 223  
Db 287 KETQCQ-----ECKEVKNGGQVFCCKSCSCVCFGGDKDKTCTAPQYDQ---332  
Qy 224 SRNVKHAIGGKECPLEKEACIVEGELLQCCPR--YSWRTSEWK--ECOVSLLELLEQQDP 281  
Db 333 -----VACSOCVPMQKPA-----DCPRPKWDKECECEPVK-----368  
Qy 282 H-----WHVT-----GPVCGGGIOTREYVCAQSVAPAAALRAKEVSPVEKALC 325  
Db 369 HDCKNGKVMDETTCQCTCPDAPVCTAGKER-----CGES-----CECKCINREPKEG-C 417  
Qy 326 VGPAPLPSQLCNIPCSITDCIVSSWSAWG-----LCIHENCHEPQGGKGRTEQRHVLME 379  
Db 418 AKPLVNNENTCKVCVPADKQNSPGGCGSGKSFNKLTCQCECDOSASKGLRWV-----471  
Qy 380 STGPAGHCPHLVSPCEDPMVCYRWLASBGIC-----FPDHGKCGLGHRIKAVQCONDRG 434  
Db 472 -----ADTKCE-----CQGMPEEG--CGKQTWISDKCKCECSPTI---TCQAPQI 513  
Qy 435 EDVS--GSLCPVPPPERKSCIEPCRMDCVLSEWTEWSSCSQSCSN--KNSDGKOT--RSR 489  
Db 514 LDNLCECKCPVNNLAQEKCKSP-----RQWTD--SKLCECSTTPATCEGCKQWCEG 565  
Qy 490 TILALAGEGKPCPPSQAQ-----EHLRNDHSCQWLHWETSP--WG-----PCSEDITL 537  
Db 566 ACQCICPGGKNGKNGKFFPKPSCECKCKNPTC-----TSQVWDADDECECKCFPKDQ 619  
Qy 538 VTALNATIG--WGEATCGVIGIOTRRVFCVSHVQVMVTKRCPDSTRPRTVPCFLPC--593  
Db 620 KPGCGDGGGOKWN--DRVCSGCGVPRPDCNTNGQIYNIT-----CAGCGGI 664  
Qy 594 -KXDCIVTAFSEWTPCPRMCOAGNATVKQSRRIIIOEAAANGGOECPDTLYEBERECEVVS 652  
Db 665 DKPSCPQQIYNWKTDCCEPNG-----MKEPVGG--CGAKTLDDEQ---706  
Qy 653 LCPYRWKPKWSPCLILPVSVWQGTGSGEACGKGLQTRAV-----694  
Db 707 -----CDCVPKPGGGTGAQKWDKTKCKCKEXEMPTGGCENKWKCDET 752  
Qy 695 -----SCISDDNPSAEMWEL-----KQNGMPLLYQECTVPCREDCTFTWASKFTF 741  
Db 753 CDCVPQKNTCIAPKVWDATKSCICVNPFPKNSPQVLKDTCCCGCNKVKAPOKFIE 812  
Qy 742 CSTNCEATKRRRLQTKSRKKEKCODSDILPLVETELCFCDDEFISQYXGNWSDCILPBG 801  
Db 813 NIDCACP-----NKKQCK-----APLWS-----DEF-----CDCVCPN- 842  
Qy 802 RREPHRGLRVQADSKEGEGRLFRVAVACS--DKNGRPVDSFC--SSSGYIOEKVPCP--857  
Db 843 -----SAGMKTCLSPKWNKVTCTCDGN--PPKPDCCPGTKQWMDKCKGCPNA 890  
Qy 858 -FDC-----KLSDWSSWGSSSSGIG---VRIRSKWLKKEPYNG---GRPCPKLDLKNQ 905  
Db 891 QTDGAGKKNDF-----TCSGCGSPGKLDCTGNTKWSAETCTCGGDVNRNG--NLKNF 944  
Qy 906 VHEAVPCYSCNQSVWVEWSSCKINNELSRJRCGGGTQSRKIRCVNTADGEGGAVDSN 965  
Db 945 NDNL--CQCECKN-----KQBMANCKSPRTWNYDTC-----KVCCKN-----ADSD 984  
Qy 966 LCNDEIPEPTQCSLMPCEVSEWGLWSKCPQ-----SCD-----PHTM 1007  
Db 985 DCVKPQIWLDDQ--CKCGCPASAQWT-----CPANKRFIEKSCCECKSPSPPIPOCK 1036  
Qy 1008 ORTRHLLRSLNSRTCAE-----DSQVQCL---LNENCFQFQYNLTWSTCQLSENAP 1059  
Db 1037 KWNEDKCVVECANVTKCEGFORWCDNQCK--CICPQVNTKCSDKQKFIKCEGCGCBTQIQ 1095

## RESULT 9

SM5A MOUSE  
ID SM5A MOUSE STANDARD; PRT; 1077 AA.  
AC Q62217;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Semaphorin 5A precursor (Semaphorin F) (Sema F).  
GN SEMA5A OR SEMAF OR SEMF.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NMRI;  
RX MEDLINE=96414430; PubMed=8817451;  
RA Adams R.H., Betz H., Puschel A.W.;  
RT "A novel class of murine semaphorins with homology to thrombospondin  
is differentially expressed during early embryogenesis.";  
RL Mech. Dev. 57:33-45 (1996).  
CC -!- FUNCTION: May act as positive axonal guidance cues.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: IN ADULT, DETECTED IN LIVER, BRAIN, KIDNEY,  
CC HEART, LUNG AND SPLEEN.  
CC -!- DEVELOPMENTAL STAGE: DIFFERENTIALLY EXPRESSED IN EMBRYONIC AND  
CC ADULT TISSUES. ITS ABUNDANCE DECREASES FROM E10 TO BIRTH.  
CC -!- SIMILARITY: Belongs to the semaphorin family.  
CC -!- SIMILARITY: Contains 1 Sema domain.  
CC -!- SIMILARITY: Contains 7 TSP type-1 domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X97817; CAA66397.1; ..  
DR MGD; MGI:107556; Sema5a.  
DR GO; GO:0016021; C:integral to membrane; IDA.  
DR GO; GO:0008046; F:axon guidance receptor activity; IDA.  
DR GO; GO:0007411; P:axon guidance; IMP.  
DR InterPro; IPR003659; Plexin-like.  
DR InterPro; IPR002165; Plexin\_repeat.  
DR InterPro; IPR001627; Sema.

InterPro; IPR000884; TSP1.  
InterPro; IPR008085; TSP\_1.  
Pfam; PF01437; PSI; 1.  
Pfam; PF01403; Sema; 1.  
Pfam; PF00090; TSP\_1; 5.  
PRINTS; PR01705; TSP1REPEAT.  
DR SMART; SMO0423; PSI; 1.  
DR SMART; SMO0630; Sema; 1.  
DR SMART; SMO0209; TSP1; 6.  
DR PROSITE; PS00092; TSP1; 6.  
Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;  
Developmental protein; Glycoprotein.  
FT SIGNAL 1 21  
FT CHAIN 22 1077  
FT DOMAIN 22 971  
FT TRANSMEM 972 992  
FT TRANSMEM 993 1077  
FT DOMAIN 993 1077  
FT DOMAIN 226 507  
FT DOMAIN 540 593  
FT DOMAIN 595 651  
FT DOMAIN 653 702  
FT DOMAIN 707 765  
FT DOMAIN 784 839  
FT DOMAIN 841 896  
FT DOMAIN 897 944  
FT CARBOHYD 147 147  
FT CARBOHYD 168 168  
FT CARBOHYD 227 227  
FT CARBOHYD 277 277  
FT CARBOHYD 323 323  
FT CARBOHYD 367 367  
FT CARBOHYD 536 536  
FT CARBOHYD 591 591  
FT CARBOHYD 717 717  
FT CARBOHYD 933 933  
SQ SEQUENCE 1077 AA; 120826 MW; EDAB0DDA42789FF CRC64;  
Query Match  
Best Local Similarity 4.3%; Score 351.5; DB 1; Length 1077;  
Matches 161; Conservative 74; Mismatches 203; Indels 255; Gaps 37;  
649 EDVSLCPYRWKPKMSCLIPVPSWVGITGSSEACGK-----GLQTRAVSCISDDN 701  
446 EEIELFPERSEPIR-SLQILHSQSVL--FVGLQEHVAKPLKRCFHOTRS-ACIGAQD 501  
702 R---SAEMMEC--LKOTNGPPLLQEC-TVPCRE---DCTFTAWSKFTPCSTNCEATKS 751  
502 PYCGWDAAWKKTSLEESLSLWQDQSIPTCTRNLTVDGSGFWSWPTPC-THDGT-- 558  
752 RRRQLTGSRKKEKQSDLYPLVETELC---PCDEFISQPYGNWSDCILPEGRREPHRG 808  
559 -----AVGSLCRSRSCDRPAPQ-CGGW----- 580  
809 LRVQADSKCEGELFRVAVACSKNGRPVDSFSCSSCYIQEKCVPICPFDPCKLSDKSSW 868  
581 -----QC-EGPRMETNCSRNGG-----WTPWTWS 604  
869 GSCSSSCGIGVIRSKWLKE-KPYNGRPPCPKLDLKNQVHEAVFCYSECNGYSWVVEH-- 925  
605 SPCSTTGIGFQVRQSCSNPTFRHGRVGVQGNREERY-----CNEHLLCPHPVF 655  
926 -----WSSCKINNELSRICGGGTQSKIRCVNTADGEGAVDSNLQNQDEIPPEQTSC 979  
656 WTGWGFWERC-----TAQCGGGIQAARRTCENGFDGAGSNVEYHPCNTN----- 699  
980 SLMCFNECVMSEGLWSKPCSPCDPHTWQRRH-----LL---RPSLSNRCAE 1026  
700 --ACELKKTITWTPTWPNVINDNGHVEQRFYRCKARLPDNLLEVGRIEWRVCS 757  
1027 DSQVQPCLLNENCFOP-----QVNL-----TEWSTQLSNAPCGQGVRRLLSCV 1072  
758 DG-TSGCSTDLGSLGFLRAGRYSAHTVNGAWSAMTWSQC-----SRDCSRGIRNR--KRV 810

## RESULT 10

SM5A\_HUMAN STANDARD; PRT; 1074 AA.  
ID SM5A\_HUMAN  
AC Q13591; O60408;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Semaphorin 5A precursor (Semaphorin F) (Sema F).  
GN SEMA5A OR SEMAF.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=98125554; PubMed=9464278;  
RA Simmons A.D., Puschel A.W., McPherson J.D., Overhauser J., Lovett M.;  
RT "Molecular cloning and mapping of human semaphorin F from the Cri-du-  
chat candidate interval."  
RL Biochem. Biophys. Res. Commun. 242:685-691(1998).  
[2]  
RN SEQUENCE OF 1-494 FROM N.A.  
RP Kalicki J., Harmon G.;  
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: May act as positive axonal guidance cues.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: Belongs to the semaphorin family.  
CC -!- SIMILARITY: Contains 1 Sema domain.  
CC -!- SIMILARITY: Contains 7 TSP type-1 domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U52840; AAC09473.1; -;  
CC EMBL; AC004615; AAC14668.1; -;  
CC PIR; JCS928; JCS928.  
CC Genew; HGNC:10736; SEMA5A.  
CC GO; GO:0007155; P:cell adhesion; TAS.  
CC GO; GO:0007267; P:cell-cell signaling; TAS.  
CC GO; GO:0007399; P:neurogenesis; TAS.  
CC InterPro; IPR003659; Flexin-like.  
CC InterPro; IPR002165; Flexin\_repeat.  
CC InterPro; IPR001627; Sema.  
CC InterPro; IPR000884; TSP1.  
CC InterPro; IPR008085; TSP\_1.  
CC Pfam; PF01437; PSI; 1.  
CC Pfam; PF01403; Sema; 1.  
CC Pfam; PF00090; TSP\_1; 6.  
CC PRINTS; PR01705; TSP1REPEAT.  
CC SMART; SMO0423; PSI; 1.  
CC SMART; SMO0630; Sema; 1.  
CC

```
DR SMART; SM00209; TSP1; 6.  
DR PROSITE; PS0092; TSP1; 6.  
KW Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;  
KW Developmental protein; Glycoprotein.  
FT SIGNAL 1 22  
FT CHAIN 23 1074  
FT DOMAIN 23 968  
FT TRANSMEM 969 989  
FT DOMAIN 990 1074  
FT CYTOPLASMIC (POTENTIAL).  
FT SEMA.  
FT DOMAIN 226 507  
FT DOMAIN 540 593  
FT DOMAIN 595 651  
FT DOMAIN 653 702  
FT DOMAIN 707 765  
FT DOMAIN 784 839  
FT DOMAIN 841 896  
FT DOMAIN 897 944  
FT CARBOHYD 142 142  
FT CARBOHYD 168 168  
FT CARBOHYD 227 227  
FT CARBOHYD 277 277  
FT CARBOHYD 323 323  
FT CARBOHYD 367 367  
FT CARBOHYD 437 437  
FT CARBOHYD 536 536  
FT CARBOHYD 591 591  
FT CARBOHYD 717 717  
FT CARBOHYD 933 933  
FT CONFLICT 56 56  
FT CONFLICT 149 149  
FT CONFLICT 382 382  
FT CONFLICT 494 494  
SQ SEQUENCE 1074 AA; 120570 MW; EE3DB763CBE29407 CRC64;  
  
Query Match.  
Best Local Similarity 4.2%; Score 345.5; DB 1; Length 1074;  
Matches 141; Conservative 59; Mismatches 168; Indels 255; Gaps 30;  
  
QY 794 SDCL-----PEGRRPHRLGLRVOADSKCGLRFRAY----- 827  
DB 441 SCLLEBIEELFPERRRPIRLSLQLHSQSVLFVGLRHVVKIFLKRQFYRTASTCIGAQ 500  
QY 828 -----ACSDKN----- 827  
DB 501 DPYCGDWMVKKCTSLSESLMTQWESISACPTRLNLTVDHGFVMSPTPCHTDGSV 560  
QY 838 DPSCSSSGVIOEKVPCP-----FDCK-----LSDHSSWGSCSSCGIGV 879  
DB 561 GSCLCRT-----RSCDSPAQCGGWQCEGPGMEIANCNRNGWTPTWTSWSPCTTCGIGF 615  
QY 880 RIRSKWLKE-KPYNGGRPCPKLDLKNQVHAFCYSECNQYVWVVEH-----WSSCK 930  
DB 616 QVRQSCSNPTPRHGRVVCVQNRERY-----CNEHLCPPHFWTGWGPWERC- 665  
QY 931 INNELSLRGGGTQSKIRCVNTADGEGAVDSNLNQDEIPPEFQSCSL-MCPNECV 989  
DB 666 -----TAQCGGSIQARRICENGPDGAG-----CN-----VEYQSCNTNCPCLKKT 707  
QY 990 SEWGLWSKQSCDPHTMQRTH-----LL-----RPSLNSRTCAEDSQVQCLLN 1036  
DB 708 TPWTPTVNVISDNGDHYEQFRYTCARLADPNLLEVGQRLEMYCYSDG-TSCGSTD 766  
QY 1037 ENCFQF-----QYNL-----TENSTCOLSENAPCGGVTRILLSVCVSGKP-VSM 1081  
DB 767 GLSGDFLRAGRYSAHTVNGAWSAWTSWSQC-----SRDCSRGIENR--KRVCNPEPKYGG 820  
QY 1082 DQCEHNLXEPQMSIPCLVECVNQLSGWTATWCSOTCGHGRMSRTRFLIMTQGE 1141  
DB 821 MCLGPLELEQECNLTCPVDGWSC-----WSPWTKSATCG-GGHVMTTRSCSNP----- 871  
QY 1142 GRPCPTELTOEKTCPVTPCYSWVLGNWSACKLEGDCGEGVQIRSLSCMVHSGSISHAAG 1201  
DB 872 -----APAY-----GGDICLGL----- 883
```

RESULT 11  
ID SMSB\_HUMAN STANDARD; PRT; 1093 AA.  
AC Q9P283;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Semaphorin 5B precursor.  
DE SEMA5B OR KIAA1445.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=20277482; PubMed=10819331;  
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.,  
RT "Prediction of the coding sequences of unidentified human genes. XVII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro."  
RL DNA Res 7:143-150(2000).  
CC -!- FUNCTION: May act as positive axonal guidance cues (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- SIMILARITY: Belongs to the semaphorin family.  
CC -!- SIMILARITY: Contains 1 Sema domain.  
CC -!- SIMILARITY: Contains 7 TSP type-1 domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB040878; BAA95969.1; ALT\_INIT.  
CC Genew; HGNC:10737; SEMA5B.  
CC InterPro; IPR003659; Plexin-like.  
CC InterPro; IPR002165; Plexin\_repeat.  
CC InterPro; IPR001627; Sema.  
CC InterPro; IPR000884; TSP1.  
CC InterPro; IPR008885; TSP\_1.  
CC Pfam; PF01437; PSI; 1.  
CC Pfam; PF01403; Sema; 1.  
CC Pfam; PF00090; tsp\_1; 5.  
CC PRINTS; PR01705; TSP1REPEAT.  
CC SMART; SM00423; PSI; 1.  
CC SMART; SM00630; Sema; 1.  
CC SMART; SM00209; TSP1; 4.  
CC PROSITE; PS50092; TSP1; 5.  
KW Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;  
KW Developmental protein; Glycoprotein.  
FT SIGNAL 1 26  
FT CHAIN 27 1093  
FT DOMAIN 20 978  
FT TRANSMEM 979 999  
FT DOMAIN 1000 1093  
FT CYTOPLASMIC (POTENTIAL).  
FT SEMA.  
FT DOMAIN 236 518  
FT DOMAIN 551 605  
FT TSP TYPE-1 1.  
FT TSP TYPE-1 2.  
FT TSP TYPE-1 3.  
FT DOMAIN 664 713





Pfam; PF02345; Tila; 25.  
Pfam; PF00094; vwd; 4.  
SMART; SM00181; EGF; 2.  
SMART; SM00274; FOLN; 11.  
SMART; SM00137; NAM; 2.  
SMART; SM00214; VWC; 17.  
SMART; SM00215; VWD; 4.  
PROSITE; PS00022; EGF\_1; 1.  
PROSITE; PS01186; EGF\_2; 18.  
PROSITE; PS00026; EGF\_3; 1.  
PROSITE; PS00740; MAM\_1; FALSE\_NEG.  
PROSITE; PS00060; MAM\_2; 3.  
Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;  
KW Repeat.

	1	17	POTENTIAL.
FT CHAIN	18	5376	ZONADHESIN.
FT DOMAIN	18	5310	EXTRACELLULAR (POTENTIAL).
FT TRANSMEM	5311	5337	POTENTIAL.
FT DOMAIN	5338	5376	CYTOSOLASMIC (POTENTIAL).
FT	MAM 1.		
FT	MAM 2.		
FT	MAM 3.		
FT	80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)		
FT	(MUCIN-LIKE DOMAIN).		
FT	VWFD 1 (PARTIAL).		
FT	VWFD 2.		
FT	VWFD 3.		
FT	VWFD 4.		
FT	VWFD 5 (PARTIAL).		
FT	VWFD 6 (PARTIAL).		
FT	VWFD 7 (PARTIAL).		
FT	VWFD 8 (PARTIAL).		
FT	VWFD 9 (PARTIAL).		
FT	VWFD 10 (PARTIAL).		
FT	VWFD 11 (PARTIAL).		
FT	VWFD 12 (PARTIAL).		
FT	VWFD 13 (PARTIAL).		
FT	VWFD 14 (PARTIAL).		
FT	VWFD 15 (PARTIAL).		
FT	VWFD 16 (PARTIAL).		
FT	VWFD 17 (PARTIAL).		
FT	VWFD 18 (PARTIAL).		
FT	VWFD 19 (PARTIAL).		
FT	VWFD 20 (PARTIAL).		
FT	VWFD 21 (PARTIAL).		
FT	VWFD 22 (PARTIAL).		
FT	VWFD 23 (PARTIAL).		
FT	VWFD 24 (PARTIAL).		
FT	VWFD 25.		
FT	EGF-Like.		
FT	BY SIMILARITY.		
FT DISULFID	5263	5274	BY SIMILARITY.
FT DISULFID	5268	5283	BY SIMILARITY.
FT DISULFID	5285	5294	BY SIMILARITY.
FT CARBOHYD	339	339	N-LINKED (GLCNAC.)
FT CARBOHYD	499	499	N-LINKED (GLCNAC.)
FT CARBOHYD	1216	1216	N-LINKED (GLCNAC.)
FT CARBOHYD	1239	1239	N-LINKED (GLCNAC.)
FT CARBOHYD	1314	1314	N-LINKED (GLCNAC.)
FT CARBOHYD	1814	1814	N-LINKED (GLCNAC.)
FT CARBOHYD	1908	1908	N-LINKED (GLCNAC.)
FT CARBOHYD	1933	1933	N-LINKED (GLCNAC.)
FT CARBOHYD	2028	2028	N-LINKED (GLCNAC.)
FT CARBOHYD	2111	2111	N-LINKED (GLCNAC.)
FT CARBOHYD	2142	2142	N-LINKED (GLCNAC.)
FT CARBOHYD	2332	2332	N-LINKED (GLCNAC.)
FT CARBOHYD	2533	2533	N-LINKED (GLCNAC.)
FT CARBOHYD	2575	2575	N-LINKED (GLCNAC.)
FT CARBOHYD	2692	2692	N-LINKED (GLCNAC.)
FT CARBOHYD	2812	2812	N-LINKED (GLCNAC.)
FT CARBOHYD	3052	3052	N-LINKED (GLCNAC.)
FT CARBOHYD	3065	3065	N-LINKED (GLCNAC.)
FT CARBOHYD	3144	3144	N-LINKED (GLCNAC.)
FT CARBOHYD	3172	3172	N-LINKED (GLCNAC.)

Query Match 4.1%; Score 335; DB 1; Length 5376;  
Best Local Similarity 20.3%; Pred. No. 5.1e-14;  
Matches 325; Conservative 154; Mismatches 523; Indels 596; Gaps 97;

	18	CEHPALQBPTEQ--ACLIPCRDCVVFSEFLPMWSNCSGKGGKQLQHRTRAVIAPPLFGGL 74	
QY	2933	CATITLQCPAHSHTNCLPPCQPSCLDSE----	GHC-----2964
DB	2965	-----EGSTTKAPSACQEGCVCEPDYVVLNNKCVPRIEGCKDAQGLVIPADKTWI--N 3016	
QY	75	QCNPNTESRACDAPISCPLG---EEFY-TFSLKVGPMSKCRPLPHLKEI-NPSGRTVLDNFN 129	
DB	2965	-----EGSTTKAPSACQEGCVCEPDYVVLNNKCVPRIEGCKDAQGLVIPADKTWI--N 3016	
QY	130	SDSNERTVTFKE---QSVKAHHHSKSWALEIGYQTRYOVSVCTSRDGO---NAWLISLCQDSF 183	
DB	3017	RGTOSTCTKRGAIQQCFQCPSETYCKDI--EDGNSCNTIRISLQCPANFNFTSLCPSCQ 3074	
QY	184	PLTVQSCI-MPKDCETSSQWSWSPCSK--TCRSGLLPGRFSRNRVKHMAIOGGKECPE 240	
DB	3075	P-----SCGNTDVHCEGSSPNLTSSCREGCVCQSG-----Y 3105	
QY	241	LLEKCACTIVEGELIQOCPRIYSWRTSWKRCQVLSLLIQDPHHVTGPVC-----GGG 293	
DB	3106	VLHNDKCI-----LRNQ-----GCKDAQGLIPEGKT--WITSQ--CTQSCNCTGGA 3149	
QY	294	IQTR-----EYVC-----AQSVFAAAALRAKEVS-----RPV---EKALCVGPAP 330	
DB	3150	IQCNFOCPFKTYCKDLKDGSNNCTNIPLQCPAHSRYTNCLPSCPFLCLDPGLCEGTSP 3209	
QY	331	-LPSQLCNIPGSTCISSWSAWGLCIHEN-----CHEPGKKGFPR-----QRH 375	
DB	3210	KVPS-----TREGICQP-----GYLMHKNKVLRIFCGKNQTQG--AFISADKTWISR 3258	
QY	376	VLWESTGPAG--HGPHLVESVPCEDPWCYRWLASGECFFDHGKCGLGHLIRAKVQN-D 432	
DB	3259	CTQSTCTCPAGAHCRNP-----KCPSG-----TYCKNGD 3287	
QY	433	REEDVSGSL--CP-----VPPPPERKSCEIFCRMDCVLSWTWSSCSQSC-----476	
DB	3288	NGSSNCTEITLQCPNTSQFTCLP-----SCVPCSNCRCVTSVPSVSSCRGCLCNHG 3341	
QY	477	-----SNKNSDGKTRSRITLALAGEGKCPPSQALQHRL--CNDHSCWQLHWET 526	
DB	3342	FVFESEDKCVPRTOCCCKDARGAIIIPAGKTWTSKGCTQSCACVEGNIQCNFQC-----3394	
QY	527	SPWGSCSEDILTVALNATIWNGBEATCGVGITRRRVFCVZSHGVQVMKRCPTSTRPETV 586	
DB	3395	PETYCKDNS-----EGSSTC-----TKTLQCPAHTQVTS 3426	
QY	587	RPCFLP-----CKKDCI-----VT 600	
DB	34		



```
PT DISULFID 665 677 BY SIMILARITY.
PT DISULFID 672 690 BY SIMILARITY.
PT DISULFID 684 699 BY SIMILARITY.
PT DISULFID 725 737 BY SIMILARITY.
PT DISULFID 732 750 BY SIMILARITY.
PT DISULFID 744 759 BY SIMILARITY.
PT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
PT NON TER 867 867
SQ SEQUENCE 867 AA; 91817 MW; 9538F2108E787B49 CRC64;

Query Match 3.98; Score 324; DB 1; Length 867;
Best Local Similarity 21.38; Pred. No. 3.6e-14;
Matches 232; Conservative 90; Mismatches 355; Indels 414; Gaps 58;

29 QACL-IPCRDVCVSEFLPNSKGC-GKKLQHRTRAVIAPLFLGGLQC-----PML 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
19 RACFAAACPDGVTWSRNSPCSEPCGGVTAHRE---CHPQNGRICATILPGGPPST 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 TETRAC-----DAPISCPLEBEYTFSLKVP-----WSKCRPLHLKEI 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
76 RETRPCQDGCNPVTCSEBELVFHACVPCPLTCDDISQATCPDRPCGGPGCWCPCAGQVL 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
119 NPSGRTVLNDSNSNERVTFKHQSYKAHHHSKWAIEIGYQTRQVSCTRSDGO----- 171
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 GAQRCV-----W-----PQCPCCL-VDSGRYWPQGR 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
172 ---NAMLSLCLQDSFPFLTVQ---SCIMPXDCETISQMSWSPSCSKTRCSGLLFGFRSR 225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
162 VKTDCQLCVC-QDGRPRRCQPSLDCAV--NCGWSAWSMAECLGPGSRSVQWSPSPN- 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 NVKHMATGGKCEPELLEK-----EACIVGELLQOCPRYSWRTSEWKECQVSLLL 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
218 --NPRPAGRHQCRGHRKARRCOTCECECEODGRVHRVGER--WRAGPCRVCO---CL 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
277 EQQDPHHVTPVCGGG-----IQRETVYCAQSVPAALAAALRAKEVRPVEKALCVGP 328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
271 HDGSAR---CSPYCPGLGCPQDVLVIEGVGSCCHCVPGGENQTVHPMATPV-----P 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
329 APIPSQLNCIPCTDCIVSWSWAGLCIHENCHEPOGKGFRTRQHRVLMESTGPAGHCP 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
321 APTFSPQIGAPLITYLLPPP-----GDCPSPLGLARL-----PEGSIF 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 HLVESVPCEDPMCVRLASEGICFPDHGKGLGHRILKAVQNDREDVSGSLCPVPPP 448
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360 --ASSQQLHP---AWNA-----ILR-----PAGAP 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 ERKSCETPCRMDCVLSEWTEWSCSCSNKNS-----DGKQTRSRITILALAGEGKP 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 -----GWSPVHADTQGHTPPPVYLQDLQPRNLGIIIVQAGS-- 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
502 CPPSQALQEHRLCNHSCMLHWETSPWGPCSEDTLVLTALNATIGNGEATCGV---GLO 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
421 ---SDMLQV-----SSDGLHWS---YRDIQGTQAPAPLPKQNWNGSTVMFARMVQ 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
559 TRRFVCKSHVGQVTKRCDSTRPRTVPCFLPKCKDCIIVTAFSEWTPCPRMCOAGNAT 618
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
469 ARHVRVWPSD-GHHQAAPSSDANLDGLPLRVLLGCE-----PAPLCILGVG--- 512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
619 VKQSRVRIIIOEAA-----NGQCECPDTLYEE----- 645
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513 -----HRCVSGEAPRGAAPCDGVEDCKGSDERGCVTPPAGAGRIESTAWSSAPSAQFG 567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
646 ---RECEDESILCPVYRWKQKSPCILVPE-----SVMGIGTGSSE 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
568 QLPPQPSGELAABEAADHWHFGRSP---VPTGKGASLGSSEPHSPGGSV-QTVTPTSQ 623
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
684 ACKGKL--QTRAVSCISDDNRSAMMECLKQTNMPLLVQECT---VPC----- 727
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
624 PEAQALAFEMAAVTVLPFHPMVTPEVAGRSTTPGPFPHVQCSPGVQVPCVILGCVELEQL 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
728 ---REDCTF---TAWSKFT-----PCSTNCEATKSRRLQLTGSKRKKERQDSLDLYF 773
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 14
AT17 HUMAN
ID AT17 HUMAN STANDARD; PRT; 1095 AA.
AC Q8TE56;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAMTS-17 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 17) (ADAM-TS 17) (ADAM-TS17).
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTS, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains."
RL Gene 283:49-62(2002).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in fetal lung, in adult brain,
CC prostate, and liver.
CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By
CC similarity).
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.
CC -!- SIMILARITY: Contains 1 PLAC domain.
CC -!- SIMILARITY: Contains 5 TSP type-1 domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AJ315735; CAC86016.1; -.
CC MEROPS; M12.027; -.
CC Genem; HGNC:17109; ADAMTS17.
CC MIM; 607511; -.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR004818; Pept_M12B.
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001590; Peptidase M12B.
CC InterPro; IPR002870; Peptidase M12B_N.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR008085; TSP_1.
```

DR Pfam; PF01562; Pep M12B proper; 1.  
 DR Pfam; PF01421; Reptolysin; 1.  
 DR Pfam; PF00090; Tsp; 1; 5.  
 DR PRINTS; PR01705; TSP1REPEAT.  
 DR SMART; SM00209; TSP1; 5.  
 DR PROSITE; PS00215; ADAM\_MBP; 1.  
 DR PROSITE; PS00546; CYSTEINE SWITCH; FALSE\_NEG.  
 DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 DR PROSITE; PS00214; DISINTEGRIN\_2; FALSE\_NEG.  
 DR PROSITE; PS00900; PLAC; 1.  
 DR PROSITE; PS00092; TSP1; 5.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix.  
 FT SIGNAL 1 27  
 FT PROPEP 28 223  
 FT CHAIN 224 1095  
 FT DOMAIN 224 452  
 FT DOMAIN 453 542  
 FT DOMAIN 543 598  
 FT DOMAIN 599 701  
 FT DOMAIN 702 779  
 FT DOMAIN 800 860  
 FT DOMAIN 861 922  
 FT DOMAIN 925 968  
 FT DOMAIN 972 1029  
 FT DOMAIN 1045 1084  
 FT DOMAIN 60 120  
 FT SITE 201 201  
 FT METAL 389 389  
 FT ACT SITE 390 390  
 FT METAL 393 393  
 FT METAL 399 399  
 FT CARBOHYD 167 167  
 FT CARBOHYD 483 483  
 FT CARBOHYD 785 785  
 FT CARBOHYD 790 790  
 FT CARBOHYD 832 832  
 FT CARBOHYD 839 839  
 FT CARBOHYD 894 894  
 SQ SEQUENCE 1055 AA; 121099 MW; A5C6540484754D5D CRC64;  
 Query Match 3.9%; Score 323; DB 1; Length 1095;  
 Best Local Similarity 21.3%; Pred. No. 5.5e-14;  
 Matches 194; Conservative 66; Mismatches 258; Indels 392; Gaps 51;  
 QY 85 CDAPISCPGEE-----YTESLKVGPWSKRLPHLKEINPSPGRTVLDNSDSNERVTFK 139  
 DB 367 CSAKKKVLADNGLNLAFTHAELG-----HNLGMHDDDD----- 402  
 QY 140 HQSYKAHH-----SKSWAIEIGYQTRQVSVCTRSQQN---AMLSLCLQDSF 183  
 DB 403 HSSCAGRSHIMSGEWNKGNFSDLSWS-----SCSRDDLENFLKSKVSTCLLVTD 452  
 QY 184 PLTVQSCIMPDKCETSQWSSPSCSKTRCSGLLPGRFSR-SRNVKMAVGGKCEPELL 242  
 DB 453 PRSQHTVLPHLKPMHYSANEQC-----QILFGMNATFCRNMELMCA----- 497  
 QY 243 EKEACIVEGELLQOCPRYSWTSWSEKQCVSLLEQQDDPHHVTGPGVGGGIQPREVYCA 302  
 DB 498 --LWCLVSGD-----TSCKTKL-----DP--PLDGTGCGA-----DKWC- 527  
 QY 303 QSVFAAALRAKEVRPVEKALCVGPAPLPQCLNIPGSTDCIVSSWSANGLCIHENCHE 362  
 DB 528 -----RAGE-----CVSKTPIPEHV-----DGDWSPWGAWSMC-SRTC-- 559  
 QY 363 PQGKGGRTRORHVLWSTGPAG-HCPHL-YESVPCDEPMCYRWLAS----- 407  
 DB 560 ---GTGARFRKCDNPPPGGTHCPGASVEHAVENLPCPKGLPFRQOQOAHRLS 616  
 QY 408 -----EGICFP-----DHGKC----- 418  
 DB 617 PKKGLLTAVVDDPKPCLYCSPLGKESPLLVADRLDVGTPGPGYETDLCVHGKQKIGC 676  
 QY 419 -GLGHRILK-----AVQOND-----RG-----EDVSGSL----- 441  
 DB 677 DGIISAAKEDRCGVCSGDGKTCHLVKGFPHAGTALKDSKGSINSIDWKIELPGBEQI 736  
 QY 442 -----CPVPPPPKRSCEIP 456  
 DB 737 AGTTVYVRRGLWEKISAKGPTKPLFLHMLVLLFFHDQDYGHIHYETVPVNRTAENQS--EP 794  
 QY 457 CRMDCVLSWTE--WSSCSQSCSNKNSDGKTRSRITLALAGEGK-----PCP-PS 505  
 DB 795 EKPQSLFIWTHSGWEGCVQGG-----GER--RTIVSCTRIVNKITLIVNDSDCPOAS 847  
 QY 506 QALQPHRLCNDSHSCMLHWETSPMGPCSEDTLVLTALNATIGMNGEATCGVGIQTRRVFCV 565  
 DB 848 RPEQVRRCNLHPC-QSRWAGPWSPCS-----ATCEKGFQHREVTCV 889  
 QY 566 -----KSHVGQVMTKRCPSDSTRPCTVPCFLPCKKDCI-VTAFSEWTPCPRMCOGNATV 619  
 DB 890 YOLQNGTHA-TRPLYCP-GPRPAVQSC-----GQDCLSIWEASEWSQCSASC--GKGVW 942  
 QY 620 KQSYRIIIEAANGGQCPDPTLYEERECEVSLCPVYRWKPKQKSPCILVPESVWQGIT 679  
 DB 943 K-----RTVACTNSQKCDASTRPRAEEACEEDVSGC--YEMKTGDWSTC----- 984  
 QY 680 GSSEACGKGLQTRAVSCISDD-----NBSAEMMECLKQTNGMPLLVQECVTP-- 726  
 DB 985 --SSTCGKGLQSRVVQCMHKVYGRHSGECPALSKPAPYRCYQEVNDRINANTITSRL 1042  
 QY 727 --CREDCFTAWSKFTPGCTNCEATKSRRLQUTGSKRKEKQSDLYPLVETELC--PC 782  
 DB 1043 AALTYKCFEDQWTVY-----CRVI-----REKNLCQDMRWY-----QRCCQTC 1080  
 QY 783 DEFISQPYGN 792  
 DB 1081 RDF-----YAN 1086  
 RESULT 15  
 AT10\_HUMAN STANDARD; PRT; 1077 AA.  
 AC Q9H324;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).  
 GN ADAMTS10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1] RN  
 RP SEQUENCE FROM N.A.  
 RA Apte S.S.;  
 RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple  
 RT thrombospondin type I repeats.";  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- COPACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 CC matrix (By similarity).  
 CC -1- DOMAIN: The spacer domain and the TSP type-1 domains are important  
 CC for a tight interaction with the extracellular matrix (By  
 CC similarity).  
 CC -1- SIMILARITY: Belongs to peptidase family M12B.  
 CC -1- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -1- SIMILARITY: Contains 1 PLAC domain.  
 CC -1- SIMILARITY: Contains 5 TSP type-1 domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
DB 671 VCGDGSACETIE-GVFPSPAGYEDVVMIPKGSVHIFQDNLNLSHLKLG----- 724
QY 240 ELLEKEACIVEG-----ELLO----- 255
DB 725 ---DOESLLEGLPGTQPHRLPLAGTTTQRLQDQDOVQSLALGPINASLIWMVLARTE 781
QY 256 -----QCPRYSWRTSEKQCVSLLEQQDPHWHVTPGVCGGGIQTREYV 300
DB 782 LPALRYRFNAPIARDSLPYSHVWYAPWTKCSAQ-----CAGGSQVQAVE 825
QY 301 CAQSVPAALRAKEVSRYVEKALCVGPALP--SQLCNI-PCSTDCIVSSWAWGLCIH 357
DB 826 CRNQLDSSA-----VAPHYCSAHSKLPRQRACNTEPCPPDWMVGNWS----- 868
QY 358 ENCHEPOGKKGFRTRQRHVLNVESTGPAGHCPHLVESVPCEDPMCYRWLASEGICFPDGHK 417
DB 869 -----LCSR-----S 873
QY 418 CGLGHRILKAVQN--DRGED--VSGSLCPVPPPPERKSCIEICRMDCVLSEWT--EWSS 471
DB 874 CDAGVRSRVVQORRVSAAEEKALDDSDACQPRPPVLEACHGP---TCP-PEWAALDWSE 929
QY 472 CSQSCSNKNSDGKQTRSRITILALAGEGKPCPPSQALQEHHR-----LCNDHSCWQLHWET 526
DB 930 CTPSC-----GPCLRHRVVLCXSADHRAILPFAHCSPAKPPATMRCNLRCPPARWA 983
QY 527 SPWGPCSEDTLVTALNATIGWNGEATCGVGIQTRRVFCVKSHVGVMTKCPDSTRPETV 586
DB 984 GEMGECS-----AQQGVQORQSVRCT-SHTGQA-SHECTEALRPPTT 1024
QY 587 RPCFLPKCKDCIVTAFSEWTPCPRMCOAGNATVKQSYRIIIOEANGGOCPCDTLYEER 646
DB 1025 QOCEAKCD-----SPTP-----GDGPECKDV----- 1046
QY 647 ECEDVSLCPV 656
DB 1047 --NKVAVCPL 1054
```

Search completed: February 24, 2004, 01:03:07  
Job time : 51.0236 secs

```
DB 671 VCGDGSACETIE-GVFPSPAGYEDVVMIPKGSVHIFQDNLNLSHLKLG----- 724
QY 240 ELLEKEACIVEG-----ELLO----- 255
DB 725 ---DOESLLEGLPGTQPHRLPLAGTTTQRLQDQDOVQSLALGPINASLIWMVLARTE 781
QY 256 -----QCPRYSWRTSEKQCVSLLEQQDPHWHVTPGVCGGGIQTREYV 300
DB 782 LPALRYRFNAPIARDSLPYSHVWYAPWTKCSAQ-----CAGGSQVQAVE 825
QY 301 CAQSVPAALRAKEVSRYVEKALCVGPALP--SQLCNI-PCSTDCIVSSWAWGLCIH 357
DB 826 CRNQLDSSA-----VAPHYCSAHSKLPRQRACNTEPCPPDWMVGNWS----- 868
QY 358 ENCHEPOGKKGFRTRQRHVLNVESTGPAGHCPHLVESVPCEDPMCYRWLASEGICFPDGHK 417
DB 869 -----LCSR-----S 873
QY 418 CGLGHRILKAVQN--DRGED--VSGSLCPVPPPPERKSCIEICRMDCVLSEWT--EWSS 471
DB 874 CDAGVRSRVVQORRVSAAEEKALDDSDACQPRPPVLEACHGP---TCP-PEWAALDWSE 929
QY 472 CSQSCSNKNSDGKQTRSRITILALAGEGKPCPPSQALQEHHR-----LCNDHSCWQLHWET 526
DB 930 CTPSC-----GPCLRHRVVLCXSADHRAILPFAHCSPAKPPATMRCNLRCPPARWA 983
QY 527 SPWGPCSEDTLVTALNATIGWNGEATCGVGIQTRRVFCVKSHVGVMTKCPDSTRPETV 586
DB 984 GEMGECS-----AQQGVQORQSVRCT-SHTGQA-SHECTEALRPPTT 1024
QY 587 RPCFLPKCKDCIVTAFSEWTPCPRMCOAGNATVKQSYRIIIOEANGGOCPCDTLYEER 646
DB 1025 QOCEAKCD-----SPTP-----GDGPECKDV----- 1046
QY 647 ECEDVSLCPV 656
DB 1047 --NKVAVCPL 1054
```

Search completed: February 24, 2004, 01:03:07  
Job time : 51.0236 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

XM protein - protein search, using sw model

Run on: February 24, 2004, 00:30:51 ; Search time 110.308 Seconds  
(without alignments)  
4190.395 Million cell updates/sec

Title: US-10-022-710-2  
Perfect score: 8241  
Sequence: 1 MVRCTQKLNRTVANEICEH.....QSTPPQKPLTAYDGLDM 1465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- SPTREMBL 25:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8227	99.8	1536	Q9C014	Q9c014 homo sapien
2	4361.5	52.9	1107	Q8BHP3	Q8bhp3 mus musculu
3	4296	52.1	1502	Q9UPZ6	Q9upz6 homo sapien
4	4108	49.8	740	Q8WUVO	Q8wuv0 homo sapien
5	1352.5	16.4	446	Q43384	Q43384 homo sapien
6	725.5	8.8	4998	11 Q8CG65	Q8cg65 mus musculu
7	720.5	8.7	5146	6 Q8SPM4	Q8spm4 bos taurus
8	712	8.6	3869	5 Q86PQ3	Q86pq3 cryptospori
9	694	8.4	2165	5 Q19791	Q19791 caenorhabdi
10	661	8.0	1637	6 Q9XSU8	Q9xsv8 bos taurus
11	555.5	8.0	4123	4 Q7S851	Q7s851 homo sapien
12	587.5	7.1	1461	5 Q8MTA8	Q8mta8 caenorhabdi
13	555.5	6.7	182	4 Q43376	Q43376 homo sapien
14	474	5.8	1089	5 Q8T3A0	Q8t3a0 ciona intes
15	470	5.7	1121	5 Q966P9	Q966p9 caenorhabdi
16	455.5	5.5	1353	5 Q9V9Q7	Q9v9q7 drosophila

Q966Q0 caenorhabdi  
Q94862 homo sapien  
Q8CDB8 mus musculu  
Q9GLX9 bos taurus  
Q90884 caenorhabdi  
Q9NCB6 homo sapien  
Q8K2Q8 mus musculu  
Q8VCC9 mus musculu  
Q8NCD7 homo sapien  
Q9W770 gallus gall  
Q42113 brachydanio  
Q9H8X0 homo sapien  
Q42114 brachydanio  
Q8MPV5 caenorhabdi  
Q8I710 caenorhabdi  
Q76840 caenorhabdi  
Q44938 haemochus  
Q23832 cryptospori  
Q96RW7 homo sapien  
Q9VAV4 drosophila  
Q96SC3 homo sapien  
Q9GRO drosophila  
Q869A0 drosophila  
Q86829 drosophila  
Q22631 caenorhabdi  
Q811B3 mus musculu  
Q23729 cryptospori  
Q94446 chironomus  
Q9V9Q9 drosophila

ALIGNMENTS

RESULT 1

Q9C014 PRELIMINARY; PRT; 1536 AA.  
ID Q9C014;  
AC Q9C014; 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein KIAA1679 (Fragment).  
GN KIAA1679.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21082932; Pubmed=11214970;  
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIX.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro."  
RL DNA Res. 7:347-355(2000).  
DR EMBL; AB051466; BAB21770.1; -.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; tsp.1; 8.  
DR SMART; SM00209; TSP1; 16.  
DR PROSITE; PS50092; TSP1; 12.  
KW Hypothetical protein.  
FT NON TER 1  
SQ SEQUENCE 1536 AA; 171242 MW; B5121C8160CEDCE CRC64;

Query Match 99.8%; Score 8227; DB 4; Length 1536;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1464; Conservative 1; Mismatches 0; Indels 2; Gaps 1;  
QY 1 MVRCTQKLNRTVANEICEHFALQPTTEQACLPDPCVWSBFLPWSNCSKGGCKLQH 60  
Db 70 MVRCTQKLNRTVANEICEHFALQPTTEQACLPDPCVWSBFLPWSNCSKGGCKLQH 129  
QY 61 RTRVIAAPPLFGGLQCPNLTESRACDAPISCLPEEBEYTFSLKVGWPWSKRLPHLKEINP 120

Db RTRAVIAPPLFGGLQCPNLTSRACDAPI SCPLGEEVTFPSLKVGPWPKRLPHLKEINP 189  
QY SGRVTLDNSNSNRVTFKHQSYKAHHHKSWAIEIGYQTRQVSCITSDGQNAWLSLCLQ 180  
Db SGRVTLDNSNSNRVTFKHQSYKAHHHKSWAIEIGYQTRQVSCITSDGQNAWLSLCLQ 249  
QY DSFPLTVQSCIMPDKQCSQWSSWSPCKTCSRSGLLPGFRSRNRNVKHAIGGKECPE 240  
Db DSFPLTVQSCIMPDKQCSQWSSWSPCKTCSRSGLLPGFRSRNRNVKHAIGGKECPE 309  
QY LLEKEACIIVEGELLQOCPRYSWRSEWKECOVSLLEQDDPHMHVTPVCGGIGTREVY 300  
Db LLEKEACIIVEGELLQOCPRYSWRSEWKECOVSLLEQDDPHMHVTPVCGGIGTREVY 369  
QY CAQSVPAALAAALRAKEVSPVEKALCVGPAPLPSQLCNIPSTDCIVSSWAGWLCIHENC 360  
Db CAQSVPAALAAALRAKEVSPVEKALCVGPAPLPSQLCNIPSTDCIVSSWAGWLCIHENC 429  
QY HEPOQKGFRTQRHVLMESTGPAHGCPLHVESVPCDDPMCYRWLASGICFPDGHKCGL 420  
Db HEPOQKGFRTQRHVLMESTGPAHGCPLHVESVPCDDPMCYRWLASGICFPDGHKCGL 489  
QY GHRILKAVCONDRGDEVSGLCPVPPPERKSCIPCRMDCVLSEWTEWSSCSQSCSNKN 480  
Db GHRILKAVCONDRGDEVSGLCPVPPPERKSCIPCRMDCVLSEWTEWSSCSQSCSNKN 549  
QY SDGKQTRRTILALAGEGKPCPPSQALQEHRLCNDHSCMQLHWETSPWGPCSEDTLVT 540  
Db SDGKQTRRTILALAGEGKPCPPSQALQEHRLCNDHSCMQLHWETSPWGPCSEDTLVT 609  
QY LNATIGWGEATCGVIGTQTRVFCVKSHGVGVMTKSCPDSTETVAPCELPCKKOCIVT 600  
Db LNATIGWGEATCGVIGTQTRVFCVKSHGVGVMTKSCPDSTETVAPCELPCKKOCIVT 669  
QY AFSEWTPCPRMCQAGNATVKQSYRIIIIEAANGQECPCDTLYEERCECDVSLCPVYRWK 660  
Db AFSEWTPCPRMCQAGNATVKQSYRIIIIEAANGQECPCDTLYEERCECDVSLCPVYRWK 729  
QY POKHSPCLLPESVWQGTGSBACGKLQTRAVSCISDDNRSAMMECLKQTNWPLLV 720  
Db POKHSPCLLPESVWQGTGSBACGKLQTRAVSCISDDNRSAMMECLKQTNWPLLV 789  
QY QECTVPCREDCTFTAWSKFTCSNCAATKRRRQLTGKSRKEKQCDSDLYPLVETELC 780  
Db QECTVPCREDCTFTAWSKFTCSNCAATKRRRQLTGKSRKEKQCDSDLYPLVETELC 849  
QY PCDEFISQPYGNWSDCILPEGRREPHRGLRVQADSKGEGRLRFAVACSDKNGRPVDP 840  
Db PCDEFISQPYGNWSDCILPEGRREPHRGLRVQADSKGEGRLRFAVACSDKNGRPVDP 909  
QY FCSGSGYIQEKCVIPCPDCKLSDWSSWGSCSSCGIGVIRSKWLKEKPYNGRGPCKL 900  
Db FCSGSGYIQEKCVIPCPDCKLSDWSSWGSCSSCGIGVIRSKWLKEKPYNGRGPCKL 969  
QY DLKN--QVHEAVPCYSECNQYSWVHWSCKINNELSLRCGGGTQSRKIRCVNTADGE 958  
Db DLKN--QVHEAVPCYSECNQYSWVHWSCKINNELSLRCGGGTQSRKIRCVNTADGE 1029  
QY GGAVDSNLNENCFQYNYLWSTQCLSENAPCGQGVTRLLSCVSDGPK 1078  
Db GGAVDSNLNENCFQYNYLWSTQCLSENAPCGQGVTRLLSCVSDGPK 1149  
QY VSMDOCEHNLKEKQRMISIPCLVFCVNCQISGWTAWTECSQTCGHGRMSRTRFIIMPT 1138  
Db VSMDOCEHNLKEKQRMISIPCLVFCVNCQISGWTAWTECSQTCGHGRMSRTRFIIMPT 1209  
QY QEGRPPCPTLTQECTCPVTPCYSWVLGNWACKLEGGDCGEGVQIRSLSCWHSQGISH 1198

Db 1210 QEGRPPCPTLTQECTCPVTPCYSWVLGNWACKLEGGDCGEGVQIRSLSCWHSQGISH 1269  
QY 1199 AGRVEDALCGEMPRODSILKOLCSVPCGCHLITENSEWSTCELTCTIDGRSFETVGRQS 1258  
Db 1270 AGRVEDALCGEMPRODSILKOLCSVPCGCHLITENSEWSTCELTCTIDGRSFETVGRQS 1329  
QY 1259 RSRFTIIQSFEENODSCPOQVLETRPCTGCKVHYTWKASLWNNNERTVWCQSDGVNVTG 1318  
Db 1330 RSRFTIIQSFEENODSCPOQVLETRPCTGCKVHYTWKASLWNNNERTVWCQSDGVNVTG 1389  
QY 1319 GCSPOARPAIQCITPACRKPFSYCTQGVGCGCKGYTEIMKSNGLDYCMKVPQSEDKK 1378  
Db 1390 GCSPOARPAIQCITPACRKPFSYCTQGVGCGCKGYTEIMKSNGLDYCMKVPQSEDKK 1449  
QY 1379 ADVKNLSGKNRPVNSKIHDIFKQWSLQPLDPDGRVKIWWYGVSGGAFIMIFLIFTSLV 1438  
Db 1450 ADVKNLSGKNRPVNSKIHDIFKQWSLQPLDPDGRVKIWWYGVSGGAFIMIFLIFTSLV 1509  
QY 1439 CKKPKPHQSTPPQKPLTLAYDGDLDL 1465  
Db 1510 CKKPKPHQSTPPQKPLTLAYDGDLDL 1536  
RESULT 2  
Q8BHP3 PRELIMINARY; PRT; 1107 AA.  
ID Q8BHP3  
AC Q8BHP3; 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Weakly similar to hypothetical protein DJ1110N13.1 in chromosome 7.  
DE D130067103RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RL 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
DR EMBL; AK051714; BAC34732.1; --  
DR MGD; MGI:2443925; D130067103RIK.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0016491; P:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR005797; Cytb\_b6\_N.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; tsp1; 7.  
DR SMART; SM00209; TSP1; 9.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
DR PROSITE; PS00092; TSP1; 10.  
KW Hypothetical protein.  
SQ SEQUENCE 1107 AA; 123640 MW; B34C708B6AA65734 CRC64;  
Query Match 52.9%; Score 4361.5; DB 11; Length 1107;  
Best Local Similarity 83.9%; Pred. NO. 0;  
Matches 759; Conservative 65; Mismatches 80; Indels 1; Gaps 1;  
QY 2 VRCIOKLNRTVAVNICEHFALQPTEQAACLI PCPDCVWSEFLPWSNCSKCGKQLQHR 61  
Db 143 VRCIOKLNRTVAVNICEHFALQPTEQAACLI PCPDCVWSEFLPWSNCSKCGKQLQHR 202  
QY 62 TRAVIAPPLFGGLQCPNLTSRACDAPI SCPLGEEVTFPSLKVGPWPKRLPHLKEINP 121  
Db 203 TRAVIAPPLFGGLQCPNLTSRACDAPI SCPLGEEVTFPSLKVGPWPKRLPHLKEINP 262  
QY 122 GRTVLDNSNSNRVTFKHQSYKAHHHKSWAIEIGYQTRQVSCITSDGQNAWLSLCLQ 181





773 RKRTLVGSKKKCKKSHLYPLIETQVPCDKYNAQVGNWSDCILPEKGVEVLLGMKV 832  
812 QADSKKEGEGIRPRAVACSKDNGRPVDPSCSSGYIQEKVCIPCPDCKLSDWSSWGC 871  
833 QCDIKKGGQGVRYAMACYDQNGELVTSRCSHGYLEAEACILIPSDCKLSEWSWRC 892  
872 SSSCGIGVIRSKWLKPKYNGRCPCKLDLKN--QVHEAVPCYSECNQYVWVEHSSC 929  
893 SKSCGSGVAVSKWLREKPYNGRCPCKLDHVNQAQVYVVPCHSDCNQYLVWTEPWSIC 952  
930 KINNELSLRCGGTQSRKIRCV-NTADGEGGAVDSNLNQDILPPTQSCSLMCPNECV 988  
953 KVTFFNRENGEGVQIRKVKCMQNTADGSEHVEDYLCDPEENPLGSRCKLPCEDCV 1012  
989 MSEGLWSKCPQSDPHTMQRTRHLLR-PSLNSRTCAEDSQVQPCILLNENCFCFOYNLT 1047  
1013 ISEGWPTQCVLPNCQSGFRQSRADPIRQPADEGRSCPNAVEKEPCLNKNKYHDYNT 1072  
1048 EWSICQLSENAPCGQVTRLLSCVCSGDKPVSMDOCEQHNLEKQPMSTPCLVECVNC 1107  
1073 DWSTCQLSEKAVCGNGIKTRMLDCVRSBGKSVDLKYCEALGLEKNQWQNTSCMVCEPVC 1132  
1108 QLSGWTAWTECSQTCGHRMSRTRFIIMPTQGEGRPCPTTELTOEKTCPVTPCVSWVLN 1167  
1133 QLSDWSPWSECSQTCGLTGKMRIRRTVTPQFGDGRPCPSLMDQSKPCVPKCYRWYQY 1192  
1168 WSACKLEGDCGEGVQIRSLSCMVHSSGISHAAGRVEDALCGEMP-----QDSILKQLC 1222  
1193 WSPQVQAOQCGEGTRTNISCVVSDGSDADFSKVPDEECADIELIIDGNKNWLEESC 1252  
1223 SVPCPGDCHLFEWSEWSTCELTCDGRSFETVGRQSRRTFIIQSFENQSCPOQVLETR 1282  
1253 SQPCGDCYLDKWSWSLQCLTCVNGEDLFGGGLQVRSRPVILQELNQHLCPEOMLETX 1312  
1283 PCTGKCHYHTWASLNNRNTVWCQRSDGVNTVGGCSPQAPAAIRQICPACRPFYSY 1342  
1313 SCYDQCVEYKWMASAWKSGSRVWCQRSDGVNTVGGCLVMSQPDADRSNCPSCQPHSY 1372  
1343 CTQGGVCGCEKGYTEIMKNSGFLDYC----MKVPGSEDKKADVKNLGKRNRPVNSKIH 1397  
1373 CSEYKTCHEGGEYEWSSNTLSQCLTILPVVVLPTMEDRGVDKT-SRAVHTQPSNP 1431  
1398 IFKG--WSLQPLDPGRVKIWIYGVSGAGFLIMIFLFTSYLVCKKP-KPHQSTPPQCKP 1454  
1432 AGRGRNFWLPFGDGRKLTWYGVGAAGAFVLLIFVSMIYLACKPKPKPQRRQNNELKP 1491  
1455 LTLAYDGDLD 1465  
1492 LTLAYDGDADM 1502

## RESULT 4

Q8WUVO PRELIMINARY; PRT; 740 AA.  
ID Q8WUVO;  
AC Q8WUVO;  
DT 01-MAR-2002 (TremBLrel. 20, Created)  
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)  
DE Hypothetical protein (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
SEQUENCE FROM N.A.  
RC TISSUE=Muscle;  
RA Strausberg R.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC019344; AAH19344.1; -  
DR EMBL; BC033125; AAH33125.1; -  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; tsp 1; 3.  
DR PROSITE; PS50092; TSP1; 7.

KW Hypothetical protein. 1  
ET NON\_TER 1  
SQ SEQUENCE 740 AA; 82199 MW; 1E7F2F10FF17794B CRC64;  
Query Match 49.8%; Score 4108; DB 4; Length 740;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 733; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 113 PHLKEINSGRTVLDFNSDSNERTVFKHQSYKAHHHSKSWAIEIGYQTRQVSCSTRSDQN 172  
DB 1 PHLKEINSGRTVLDFNSDSNERTVFKHQSYKAHHHSKSWAIEIGYQTRQVSCSTRSDQN 60  
QY 173 AMLSICLDQSPPLTVQSCIMPXKDCETSWSSWSPCCKTCSRSGSLLPGRSRNRVVKHAI 232  
DB 61 AMLSICLDQSPPLTVQSCIMPXKDCETSWSSWSPCCKTCSRSGSLLPGRSRNRVVKHAI 120  
QY 233 GGGKCEPCLLEKEACIVGELLQQCPRYSWRTSWKECOVSLLEQQDPHWHVTPVCGG 292  
DB 121 GGGKCEPCLLEKEACIVGELLQQCPRYSWRTSWKECOVSLLEQQDPHWHVTPVCGG 180  
QY 293 GIQTRVYCAQSPVAAAALRAKEVSRPVEKALCVGPAPLPQCLNIPGSTDCIVSSWSAW 352  
DB 181 GIQTRVYCAQSPVAAAALRAKEVSRPVEKALCVGPAPLPQCLNIPGSTDCIVSSWSAW 240  
QY 353 GLCIHENCHEPGKGFRTORHVLWESTGPAHCPHLVSPCEDPMCYRWLASGICF 412  
DB 241 GLCIHENCHDPOGKGFRTORHVLWESTGPAHCPHLVSPCEDPMCYRWLASGICF 300  
QY 413 PDHGKCGLGHRLKAVCQNDRGEDVSGSLCPVPPPPPERKSCIEICRMDCVLSEWTESSC 472  
DB 301 PDHGKCGLGHRLKAVCQNDRGEDVSGSLCPVPPPPPERKSCIEICRMDCVLSEWTESSC 360  
QY 473 SOSCKNKSNGKQTSRTILALAGEGKPCPSQALQEHRLCNDHSCMOLHWETSPWGPC 532  
DB 361 SOSCKNKSNGKQTSRTILALAGEGKPCPSQALQEHRLCNDHSCMOLHWETSPWGPC 420  
QY 533 SEDTLVTALNATIGNGBATCGVGITQTRVFCVSHVQVMTKRCPCDSTRPETVRPCLP 592  
DB 421 SEDTLVTALNATIGNGBATCGVGITQTRVFCVSHVQVMTKRCPCDSTRPETVRPCLP 480  
QY 593 CKKQCIIVAFSEWTPCRMCQAGNATVQSYRIIIOEAANGGQECPTLYEEBCEDEV 652  
DB 481 CKKQCIIVAFSEWTPCRMCQAGNATVQSYRIIIOEAANGGQECPTLYEEBCEDEV 540  
QY 653 LCPVYRWKPKQWSPCLLVPSVWQGITGSSBACGKLOTRAVSICISDNRSAEMWELKQ 712  
DB 541 LCPVYRWKPKQWSPCLLVPSVWQGITGSSBACGKLOTRAVSICISDNRSAEMWELKQ 600  
QY 713 TNGMPLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSRRLTQKSRKKEKQDSDLY 772  
DB 601 TNGMPLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSRRLTQKSRKKEKQDSDLY 660  
QY 773 PLVETELCPCEDEFISQPYGNWSDCILPEGRPHRGLRVQADSKCEGGLFRFVACSDK 832  
DB 661 PLVETELCPCEDEFISQPYGNWSDCILPEGRPHRGLRVQADSKCEGGLFRFVACSDK 720  
QY 833 NGRPVDPSFCSSG 846  
DB 721 NGRPVDPSFCSSG 734

## RESULT 5

O43384  
ID O43384 PRELIMINARY; PRT; 446 AA.  
AC O43384;  
DT 01-JUN-1998 (TremBLrel. 06, Created)  
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)  
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
DE Hypothetical protein GS164B05.1 in chromosome 7 (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
X NCBI_TaxID=9606;
UN [1]
IN SEQUENCE FROM N.A.
RC TISSUE=Lymphoblast;
RA Andrews S., Holmes A., Nguyen C.;
AL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
JR EMBL; AC004160; AAC03417.1; -
JR InterPro; IPR000884; TSP1.
JR Pfam; PF00090; TSP_1; 3.
JR SMART; SM00209; TSP1; 7.
RW PROSITE; PS00092; TSP1; 3.
GW Hydrothermal protein.
AT NON_TER 1
TT NON_TER 446
TQ SEQUENCE 446 AA; 50444 MW; 0FC13216E84838B9 CRC64;

Query Match 16.4%; Score 1352.5; DB 4; Length 446;
Best Local Similarity 51.0%; Pred. No. 1.2e-119;
Matches 227; Conservative 78; Mismatches 133; Indels 7; Gaps 3;

Y 847 YIOEKVIPCPCDKLSDWSNGSCSSCGIGVIRIRSKWLKEPYNGRPPCKLDLKNQV 906
b 1 YIEEACIIPCPCDKLSDWSNGSRCSKSGGVKVRKWLREKPYNGRPPCKLDHVNQV 60
Y 907 HEAVPCYSCNOYSWVHEHWSCKINNELSLRCGGTQSRKIRCV-NTADGEGGAVDSN 965
b 61 YEVVPCHSQCNQLWTEPWSICKVTFVNNRENGEGVQTRKVRQCNQNTADGPSHEVDY 120
Y 966 LCNQDRIPPETOSCLMCPNECVSEWGLWSKPOSCDPHTMQRTHLRL-PSLNSRTC 1024
b 121 LCDPEEMPLGRCKLPCPEDCVISEWGPWTCVLPNCQSSFRQSRADPIRQPADEGRSC 180
Y 1025 AEDSQVQCLINENCFOFQYNLTENSTCOLSENAPCGQVTRLLSCVSDCKPVMQOC 1084
b 181 PNAVEKEPNLKNKCNHYDYNVDWSTCOLSEKAVCGNGIKTRMLDCVRSDDGSKVDLYC 240
Y 1085 EOHNEKFORMSIPICVECVNQLSGWTAWTECSQTCGHGGRMSRTRFIINFTQEGGRP 1144
b 241 EALGLEKNQWNTSCNVECPVNCQSDWSPWSECSTCGLTGKMIRRTVTFQFGDGRP 300
Y 1145 CPTLTQKTCPTVPCYSVVLGNWSACKLEGGDGGEGVQIRLSQNVHSGSTSHAAGRV 1204
b 301 CPSLMDQSKPCPKVPCYRQYQGWSPQEAQCGQGGRTNRISCVVSDGSDADFESKVD 360
Y 1205 DALCGEMPF-----QDSILKOLCSVPCPGDCHLTWSEWSTCELTICDGRSFTVGROR 1259
b 361 EFCADIELIIDGNKNVLEESCSQCPGDCYLDKWSWSLQCLICVNGEDLFGGIQVR 420
Y 1260 SRTFIQSFENQSCPQQVLETRPC 1284
b 421 SRPVIIQELNQHLCPEQMLETKSC 445

RESULT 6
28CG65 PRELIMINARY; PRT; 4998 AA.
AC Q8CG65;
AC Q8CG65;
JT 01-MAR-2003 (T-EMBLrel. 23, Created)
JT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
JT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE SCO-Spondin.
EN SCO-Spondin.
SC Mus musculus (Mouse).
SC Fukuyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
SC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
UN [1]
IN SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RA Goncalves N., Simon-Chazottes D., Creveaux I., Meiniel A.,
RA Guenet J.-L., Meiniel R.;
RT "Characterization, spatio-temporal expression and chromosomal
RT assignment of mouse SCO-spondin.";
```

```
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ491857; CAD42654.1; -
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005507; F:copper ion binding; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR000923; BlueCu_1.
DR InterPro; IPR006207; Cys knot_C.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR004421; FAS8_C.
DR InterPro; IPR001545; Gly HormoneB.
DR InterPro; IPR002172; LDL receptor_A.
DR InterPro; IPR002919; TIL Cysrich.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR008085; TSP_1.
DR InterPro; IPR005552; VC_out.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR Pfam; PF00057; ldl_recept_a; 10.
DR Pfam; PF01826; TIL_10.
DR Pfam; PF00090; TSP_1; 25.
DR Pfam; PF00093; vwc_1.
DR Pfam; PF00094; vwd; 3.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PRINTS; PR01705; TSP1REPEAT.
DR SMART; SM00192; LDLA; 10.
DR SMART; SM00209; TSP1; 25.
DR SMART; SM00214; VMC; 4.
DR SMART; SM00215; VMC_out; 9.
DR SMART; SM00216; VWD; 2.
DR PROSITE; PS00196; COPPER_BLUE; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS00022; FAS8C_3; 1.
DR PROSITE; PS00261; GLYO_HORMONE_BETA_1; 2.
DR PROSITE; PS01203; LDLRA_1; 8.
DR PROSITE; PS00068; LDLRA_2; 10.
DR PROSITE; PS00092; TSP1; 25.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 2.
SQ SEQUENCE 4998 AA; 535019 MW; DA2ABA8DA47DF225 CRC64;

Query Match 8.8%; Score 725.5; DB 11; Length 4998;
Best Local Similarity 22.3%; Pred. No. 1.4e-58;
Matches 377; Conservative 137; Mismatches 478; Indels 695; Gaps 103;

QY 18 CEHFALQPTTEQAQLIPCRDVCVVSEPLPWSN---CSKGGCKLQHRTRAVIAP-PLFGG 73
Db 3224 CNHCTMGEGRINCTDLPCQ---VSGDKCPWSKWTACSQPCRQTRTSRACVCPAQHGG 3280
QY 74 LQCPNLT-----ESRACDAPISCPGLBEEYFTSLKVGWPKC-----RLPHLK 116
Db 3281 SPCEESGGTGVQHQMEACFNATACPVDGAWSPW---GWSSCDACLGOSYRSVCSHP 3336
QY 117 EINFSGRTVLDFNSDSNERTVTFKHQSYKAAHHSKSWAIEIGYOTROVSCYTRSDQNAMLS 176
Db 3337 PISDGGKPCIG-----GYQQR-PCRNSS-----T 3360
QY 177 LCL-----QDSFFLTVQSCIMP--KDC-ETSONSSWSPCKTCRSGSLLP----- 218
Db 3361 LCTDCGGGQDLLP-----CQPCRHSQDLISLSTCQPSAGCGGCGPQGLSQHGLC 3415
QY 219 -----GFRSSSRNV---KMAIGGKCEPILLEK-----EACIVEGILQ----- 255
Db 3416 VFPVDCCHFPQPRAMGIPENRSRSVGTLSWSLEPGEVVTGPCDNCTCAVILQCHEV 3475
QY 256 -QCPR-----YSWRTSEWKECCQVSLLEQQDPHHVHTGPVCGGGIQTREVVYCAQSVFAAA 310
Db 3476 PSCFGPGIWSSW--GPWEKCSVS-----CGGGEQ----- 3502
```



SMART; SM00068; GHB; 1.  
SMART; SM00192; LDLA; 10.  
SMART; SM00209; TSP1; 25.  
SMART; SM00216; VWD; 3.  
PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
PROSITE; PS01235; CTCK\_2; 1.  
PROSITE; PS01285; FAS8C; 1.  
PROSITE; PS01286; FAS8C; 1.  
PROSITE; PS00022; FAS8C; 1.  
PROSITE; PS00261; GLYCO\_HORMONE\_BETA\_1; 2.  
PROSITE; PS01209; LDLRA\_1; 7.  
PROSITE; PS00068; LDLRA\_2; 9.  
PROSITE; PS00092; TSP1; 25.  
PROSITE; PS01208; VWFC; 1.  
PROSITE; PS01208; VWFC; 1.  
PROSITE; PS00184; VWFC; 2.  
SEQUENCE 5146 AA; 543576 MW; 724C5FB872E13DA CRC64;  
Query Match 8.7%; Score 720.5; DB 6; Length 5146;  
Best Local Similarity 21.3%; Pred. No. 4.3e-58;  
Matches 423; Conservative 155; Mismatches 571; Indels 835; Gaps 115;  
2Y 28 EQACLP-CPDRCVVEFLPMSNCKGCKGKQHRTRAVIAPLFG----- 72  
Db 3221 EEPCLLPEDRAGCWPMWSSCSRSRACQPPQGLGDCYCEGPRAQAA 3280  
2Y 73 -GLOCP--NLT-----ESRACDAPISCPGEEHYTSL-KVGMKSKRLPHLKEINPSG 122  
Db 3281 CQALPCPVNCTAIEGAIVSACGPP--CPRSCDDLVHCVHQPCCYC-----PPG 3329  
2Y 123 RTVLDNFSNERNVTPKHOSY-----KAHHSKSWAIEIGYQTRVSCSTRSDGNAMLS 176  
Db 3330 QVL---SADGTVHVQGHSCDLLLGERHHPGAQLAKPDGCN---YCTSEGLTCTD 3382  
2Y 177 LCLQDSFFLTVQSCIMPCKCETSQNSWSKSCXTCRSGSLLPGFRSRNV-----KEMA 231  
Db 3383 L-----PCVPGWACP-----WSEWATACQPCQGT-----RTRACACAPAQH-- 3422  
2Y 232 IGGKECP-----ELLKEACIVSGELLQOCP-RYSWRT-SEWKECOVSL--LLEOOD 280  
Db 3423 --GGAPCPGEAGEAGHQRETCASTPE---CPVDGANSWPMSPSPCEVCLGRSHRSRE 3476  
2Y 281 PHHVTG---PVCGGGIQTR-----VYCAQSVPAALAAKAEVSRP 319  
Db 3477 CSWPP--SEGRPCPGGHRQSRPCQGNSTQCTCAGGQDLPFCQPCPRS---CEDLSPG 3532  
2Y 320 VE-----KALCVGPAPL----- 331  
Db 3533 VEQPDSMGQPCPCGPGQLSQDGLCVTPSCRCQYQPGAMGFPENQSRAGSGLSW 3592  
2Y 332 ---PSQLCNIPC-STDCI-----VSSWSAWG-----L 354  
Db 3593 ESLEPGEVVTGPDNCTCVAGILQCOEVPACSGGLGWSGWPWEDCSVSCGGEQLRFR 3652  
2Y 355 CIHENCHEPGKGFTRQHVLMESTGAGH-----CPHLVESVPCEDPM 400  
Db 3653 CPAPCPGPARQS--RTCTQVCREAGCFAGRLRECPQSEGCPCFAHTVQGVCFAG 3710  
2Y 401 CYRWLASGICFPDGHKGLGHLRLKXAVQND-----RGEDVSGSLC--- 442  
Db 3711 C-----EEGCHPE-----GTFHRSACVQECPCVLTALWQLGGAAGADPGAHLSVLG 3759  
2Y 443 ---PVPPPE-----RKSC-----ETPCRMDCV-----LSEWTESSCSQCS 477  
Db 3760 ENGQPLGPDGLSGOSLATGCHNCSCAHGLKSCSVSEACSKAAGFSWPMGPMGCSRSG 3819  
2Y 478 NKNSDGQTRSR--TILALAGEGKPC-PPSQALQEHRLCNDHSCMLHWET----- 526  
Db 3820 GL---GTRRESQCVPMAPAGQCGCHGHWDLE---YCPSEPCGAAGSTAEPAATGLPG 3873  
2Y 527 ---SPWGPCSEDTLVTALNATIGNGEATCGVGTQTRRVFCVSHGVQVMVKCPDST 581  
Db 3874 GWLWSPWSPSCGTCTDPAHPA---W-----RSRRLCLANCTGGAASQE----- 3915

QY 582 RPETVRECFILP-----CKK-DCIIVTAFSEWTPCPRMCOAGNATVKQSRVRII 628  
Db 3916 -----RPNLPSCCTELPLCPGPGCEAGNCSTAWAPWPEPCSRGCVG-----QQRLLR-AY 3965  
QY 629 QEAANGQEGED--TLYEERECEDEVSLCPV-----YWKPKQM----- 664  
Db 3966 HPPGPGHWCVPDLVTAQERRFCNLRACVPVGGWSRSPMSWCDRSCGGRSLSRSCSS 4025  
QY 665 -----SPCI-----LVPSVMGIGTSGSEACGKGLQTRAVSCIS----- 698  
Db 4026 PPPKNGGAPCVGERHARLCPN-----TPCEGCPAGME--VWSCANRPRRCSDLQE 4076  
QY 699 -----DNRSANMEC-----LKQING-MPLLVOECT----- 724  
Db 4077 GIVQEDQACQCGCRCEPESLEQDGGCVPLGHCECTDAQHSWAPGSHQZACNNCTCRA 4136  
QY 725 -----VPC--REDCTFTAWSKFTPCSTNCEATKSRRLQLTGKSRKKEKQSDLYPLV 775  
Db 4137 QQLSCTAQCPCPPPAHCAWSRWSAWSPCSRSCGPAQQSR----- 4175  
QY 776 ETELCPDEFTISQPYGNWSDCILLPEGRREPHRGLRVQADSKEG-----EGLRFA 826  
Db 4176 -----FRSSTSGWA-----PECREE-----QSQPCQSPCPCLCLQGTFRS 4215  
QY 827 VA-----QSDKNRPPVDPSPFCSSSGVIOEKVPCPFDCKLSDWSSWSSCSG 876  
Db 4216 LGDSWLODGCQCS-----CTPEGLICEDA--ECAGLWMTFWSWDCPVSCG 4262  
QY 877 IGVIRSK-WLKEKPYNGRCPKLDLKNQVHBAVPCYSECNQYSVWVHWSCKINNEL 935  
Db 4263 GGNQVRTRVCVASAPPRGSGFCLGPDVQSQRCLWPCPALPDTCSW--GPWGPC----- 4314  
QY 936 RSLRCGGTOSRKIRC-VNTADGEGGAVDSNLNQDDEI PPETQSC-SLMCPNCVMSEWG 993  
Db 4315 -SRSCGGLASRSASCCLLAEPAP-----CNSTSPRLDTQACVAGPCLCEECVWSWS 4367  
QY 994 LWSKPOSCDPTHMQRTRHLLRPLNSR-----TCAE-DSQVQPCLLNENC-----F 1040  
Db 4368 SWTRC--SCEV-LVQQRVH-QRPAPGAGAGPPCTFLDGHFRPCLTG-NCSEDSACPPF 4422  
QY 1041 QFO-----YNTEWSTQLSNAFCGQGVTRLLSCVCS-----GKPVMSDQ 1083  
Db 4423 EFQAGSFPCTGLCATYLSPW-LCO--DLPECQPG-----CYCEGLLEQAGCGVPEQ 4472  
QY 1084 CE-QH-----NLEKPRMSIPCLVECVNQ-----LSGWTATEC 1118  
Db 4473 CNCHVSGEGAVTLAPGDRLOLGC-KEC--EQRGELOCTSGCQGLLPLSGWSEWSPC 4529  
QY 1119 -----SOTCG-----HGRMSRTRFIMPTQGE---GRP- 1144  
Db 4530 GPCLPLGLLAPASRAALEERWPDQTAGLSPTSAPTLASEQHRRLCLDPTGRPWAGDPD 4589  
QY 1145 -CPTELTOEKTCPVTP-----CYSWVLGNWSACKL-----EGGDC-GE 1180  
Db 4590 LCTVPLSQQLRCP-DPGACQDLQWGPWAGWSPCQVPCSGGFLRWREAGIPFGGCRGP 4648  
QY 1181 GVQIRLSMWHSGSI SHAAGRVEDALC----- 1210  
Db 4649 WAQTESCNMGPCGESCEAQTVPDPCANQCPRSCVDLWDRVECLQGPCRPGCRCPGQ 4708  
QY 1211 MPFQDS-----ILKQLCSVPCG-----DCH-----LTEW 1235  
Db 4709 L-VQDGHCVFSSCRCLPSPFNASWALAPAEVVRDLDCRNTCVNGSLACSHSCPTLGPW 4767  
QY 1236 SEMWTCBLTICDGRSFETVGRQSR-----RTFIIQSPENQDSC-----POOVL 1279  
Db 4768 SAMENCAPCGGTT-----KBRHSCKEGPGVTFCQAQDMEQQQDCNLQPCPCPGQVL 4822  
QY 1280 ET-----RPCT-----GKCC-----VHYTWKAS 1297  
Db 4823 SACAVSFRICSHLQPGTFCWQBPQOLGCDPCPRQLHNGTCVPPAECPCTQLSLWGLT 4882  
QY 1298 LWNNER-----TVMQCRSDGVNVTGG-----CSP-----QARPAAIRQCI 1335

```

Db 4883 LTLEEQHRELPPGTLTQNTCHCICOGGAFSCSLTDCQBCPPGETWQVAPGELGCEQT 4942
2Y 1336 CRKPFSTCTGGV-----CGCEKGYTEIMKNGELDYCMKV-----PGSE 1375
Db 4943 CREPNATETGNCGRQAPGCVQGRHFRSQEGP-----CVPVDLCCEWHGPRHPHPPGSE 4997
2Y 1376 DKKA 1379
Db 4998 WOKA 5001

RESULT 8
Q86PQ3
ID Q86PQ3 PRELIMINARY; PRT; 3869 AA.
AC Q86PQ3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TSP1 domain-containing protein TSP2 precursor.
OS Cryptosporidium parvum
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
[1]
SEQUENCE FROM N.A.
MEDLINE=22326344; PubMed=12438378;
Deng M., Templeton T.J., London N.R., Bauer C., Schroeder A.A.,
Abrahamson M.S.;
"Cryptosporidium parvum genes containing thrombospondin type 1
domains.";
RL Infect. Immun. 70:6987-6995(2002).
DR EMBL; AY190984; AAC39046.1;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0030354; P:cell differentiation; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001450; 4Fe4S ferredoxin.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR003582; Shkt_dom.
DR InterPro; IPR00436; Sushi_SCR_CCP.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00090; tsp.1; 10.
DR SMART; SM00004; NL; 6.
DR SMART; SM00254; Shkt; 5.
DR SMART; SM00209; TSP1; 11.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 3.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS50092; TSP1; 10.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 3869 AA; 429510 MW; 1F4851B7B0787B87 CRC64;

Query Match 8.6%; Score 712; DB 5; Length 3869;
Best Local Similarity 22.8%; Pred. No. 1.9e-57;
Matches 346; Conservative 153; Mismatches 448; Indels 570; Gaps 81;

Qy 33 IPCPRDCVSEFFPWSNCSCGCKGLQHRTRAVIAPPLFGL-QCPNLTESRAC-----DAP 88
Db 187 IPCPLSCTVSENGWNSRCSLTGIGHQWRERSVIKAPKDQNLFCQCPETQIRECIQDTC 246
Qy 89 ISCPGLGEEYTFSLKVPWCKRLPHLKEINPSGRVLDPNSSNSRVRTFKQSYFAHH 148
Db 247 SNCTLGFKFKAISVSGSP-----CLM-----EKGCVERREILHPFPK----- 284
Qy 149 SKSWAIEIGYQTRVSCSTRSDGQNALSLCLQDSFFLTQSCIMPXDCETSONSWSPCS 208
Db 285 -----ELGNETCKVERHSD-----CLGLHL-----GCAGP--CFEFSWGSLSLCS 323
Qy 209 KTCRSGSLLPGF-----RSRNRVKNHMAIGGKCEPPELLEKEACIVEG 251

```

```

Db 324 YYKNEGL---GFDDEASKYLLKKRRVAVARRMDA-----NCEDELEYKECDESE 371
Qy 252 ---ELLQOCPRYGRWRTSEWKECQVSLLLQQDPHWHVTGPVCGGIGTREVYCAQSVPA 307
Db 372 RHGVREALVRNMGMW---SSWSSCSTS-----CGLSRRLARWLLSEPN 413
Qy 308 AALRAKEVSRPEKALCVGPAPLPQOLCN-IPGSTD-CIVSSWSAWGLCIHENCHPEQG 365
Db 414 EEALSENLIN---SESYC-GPL-FQKACNDRSLTEGCKVSDMGOWSAGSSKSNFP-- 466
Qy 366 KKGFRTRQRHVMESTGPAGHCPHVESVPCDDPMCYRWLASSEGICFPD--GKCG 419
Db 467 --GLSKQRSVI--SLPRKGDGCPVLESRDG-----LGAC--DHSVSVSKNCNM 509
Qy 420 LGHRILKAVQND---RGEDVSGSLCPVPPPPPP-----KS 452
Db 510 FGDWSSWSTCQDDCYTRYNSSGANRSEKPKRYRHKMIIFNPKDKCKSENYEVEN 569
Qy 453 CEIICRMD---CVLSEWTESSCSQCSNKNKSGKQTRSRITILALAGEGKP----- 501
Db 570 CEGGCENANVEICQVSEWGPWSSCSANCDGSSRRIRERINSRGVKGAGSRPLTRGVSS 629
Qy 502 -CPPSQALQEHRLCNHSCMLHWETSPW-----GPCSEDTLVTALNATIG 546
Db 630 SCP---SLLEVEKCNTHPC-EYSCELSPWYSNKSSTGPDQRRGNRSDAMLEDERNEDIV 685
Qy 547 WNGEAT-CGVGI--QTRRV-----FCVYSH--VGQVTKRCPDSTRPRTVPCPLCKKDC 597
Db 686 ISDCSTKCGIGCVERSRRLISGGFVSSNEEQNYTSKVG------FLX--- 729
Qy 598 IVTAFSEWTPCPRCMQAGNATVKQSRVRIIIEAANGQCEPCPTLYEERECEDVSL-CPV 656
Db 730 -----DTIKCVRISEGCSDVQCV 747
Qy 657 YRWPKQKSPILVPESVMQGITSSAGCKGLQTRAVSCISDDNRSAEMECULKQTNMG 716
Db 748 GEW---SSWSAC-----SKSCGEGFQ-----NRSRELL-----V 773
Qy 717 PLLVQECTV-----PCREDCTFTAWSKFTPCSTNCEATKSRRLQITGSKRKKEK 765
Db 774 PSLGRKELATEELRECFERPCPSSCQVSPSEWTMC----- 810
Qy 766 QDSDLVPLVETELCPDCEFTSQPYGNWSDCILPEGRREPHRGLRVOADSKGEGGLRFR 825
Db 811 -----LGGCDE---KPY-----KRRERKILEAPAGQNC----- 836
Qy 826 AVACSDKNGRPVDPSPFCSSSGYIOEKC--VIPPFDCKLSDWSSWGSCSSCGIGVIRS 883
Db 837 -LALEEK-----VD-----AEECPRIERCPDCKVGEWSSWSECDACKGIGVEKRL 882
Qy 884 KWLKEKPYNGGRPCPKL-DLKNQVHEAVPCYSECNQYSWYVHEWSSCKINNELRSLRCGG 942
Db 883 RKVVKRESKGGAPCPNLEDLRPCSREA--CKSDCVLGEWT--EWGVC-----SKSCGG 931
Qy 943 GTQSRKTRCVNTADGEGAVDSNLNQDEIIPPETQSCS---LMCPNECVMSWGLASKCP 999
Db 932 GARSR-VREVISQNEG-----KECEFLXKEFEPNCFQCIATRDCEVQGSWWSPCS 982
Qy 1000 QSCDPHTMQRRTRHLRLRPSLNSRTCAEDSQVQCLLNENCFQFQYNLTEWSTCOLSENAP 1059
Db 983 ASCG-GGVKRGREVOVPATGGGRG-----EFELN----- 1011
Qy 1060 CGQVTRLLSCVSDSKPKVSMDOCEQHNLEKFORMS-IPC-LVECVNQLSGWTAWTE 1117
Db 1012 --QKVGNCGFKC--PGEF-----C-----IDRPAQAEVWPVCSILKAMFGCQ----- 1048
Qy 1118 CSQCGHGRMSRTRFTIMPTQGGREPCPTTELTEKT---CPVTPCYSVWLGWSACKL 1173
Db 1049 -----KRLIDVAKSNGVYPDDRPPEARIMDGCAT-CGMCV----- 1084
Qy 1174 EGGDCGEGVQIRLSJSCMVHSGSIHAAAGRVEDALCGEMPFQDSILKQLCSVPFGDCHLT 1233

```

```

1085 ---ECAPGQQLRDVGNLSCDPACNNAACRFDGDC---KNSPOKASCLPLKLEAF- 1136
1234 EWSWETCBELTIDGSRFFETVGRQSR-----RTIIOSFENQDSCPOQV 1278
1137 -FLSKKEDQVRKNGSDFLIEIGNSNGKKEAQNILKSSDNNLVVLEBSFTSQSLNS-- 1193
1279 LETRPTGCKGVHYTWKASLWN-----NNERTVWCQSDGWNVTGGCSPQAR 1325
1194 -----SSGKSNNGYFAGNQAQVQVHEMFSQEDYNNR-----HNGDGSN----- 1233
1326 PAAIROCI PACRPFYCTQGGY-CGCEKG-----YTEI-----MKNGFLDYCMKVP 1373
1234 ---TIRE-----YKGGDTVYICRDKGRKRYTAIPFYKQCKESGFQLI----- 1276
1374 SEDKADVNLSGKRRP 1390
1277 -BFGAPLLYVDGIGP 1292

RESULT 9
119791
119791 PRELIMINARY; PRT; 2165 AA.
119791 Q27524;
119791 Q1-NOV-1996 (TrEMBLrel. 01, Created)
119791 Q1-NOV-1998 (TrEMBLrel. 08, Last sequence update)
119791 Q1-OCT-2003 (TrEMBLrel. 25, Last annotation update)
119791 F25H8.3 protein.
119791 Caenorhabditis elegans.
119791 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
119791 Rhabditidae; Peloderinae; Caenorhabditis.
119791 NCBI_TaxID=6239;
119791 [1]
119791 SEQUENCE FROM N.A.
119791 CAJADSTY S;
119791 Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
119791 [2]
119791 SEQUENCE FROM N.A.
119791 MEDLINE=94150718; PubMed=7906398;
119791 Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
119791 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
119791 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
119791 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
119791 Jones M., Kersey J., Kirkstead J., Laister N., Maitreille P.,
119791 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
119791 Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showlken R.,
119791 Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
119791 Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
119791 Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,
119791 "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
119791 elegans."
119791 Nature 368:32-38 (1994).
119791 EMBL; Z69361; CAA93288.1;
119791 EMBL; Z69360; CAA93288.1; JOINED.
119791 EMBL; Z69360; CAA93287.1;
119791 EMBL; Z69361; CAA93287.1; JOINED.
119791 PIR; T21371; T21371.
119791 HSP; P15167; IDTH.
119791 MEROPS; M12.135; -.
119791 WormPep; F25H8.3; CE05729.
119791 GO; GO:004222; F:metallopeptidase activity; IEA.
119791 GO; GO:0004270; F:zinc ion binding; IEA.
119791 GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
119791 InterPro; IPR001590; Peptidase_M12B.
119791 InterPro; IPR006025; Pept_M_Zn_BS.
119791 InterPro; IPR000884; TSPI.
119791 Pfam; PF01421; Reprolysin; 1.
119791 Pfam; PF00090; tsp 1; 14.
119791 SMART; SM00209; TSPI; 18.
119791 PROSITE; PS0215; ADAM_MPRO; 1.
119791 PROSITE; PS00092; TSPI; 15.
119791 PROSITE; PS00142; ZINC_PROTEASE; 1.
119791 SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;

```

```

Query Match      8.4%; Score 694; DB 5; Length 2165;
Best Local Similarity 22.9%; Pred. No. 4.5e-56;
Matches 351; Conservative 133; Mismatches 556; Indels 492; Gaps 84;

QY 18 CEHPALQPTPEQAACLPAPR-----DCVSEFTLWNCSCGCKKQHRTRAVIAP-PLF 71
DB 586 CHH-----GACVRLAPESLTKIDQWGDWRSWGECSTCGGVQKGLRDCDCKPRN 637
QY 72 GGLQCPNLTS-RACDAPISCPLEGEETPSLVKVPWPKRLPHELKEINPSGRVTLDFNS 130
DB 638 GGYKVCQQRERYASCTNQ-ECFNDTQPY-----REVQCS-----EFN- 673
QY 131 DSNERTVFKHQSKAHHSKSWAIEIGYQTRQVSCSTRSDGQNALSL-----CLQD 181
DB 674 --NKDIGIQGVASTNTTHWPKYANVAPNERCKLYC-RLSGSAAFYLLRDKVVDGTCDRN 730
QY 182 SFPLTVQSCIMPKDC-----TSQWSSWSPCKTCSRSGLLPG----- 219
DB 731 GDDICVAGACYPAGCDQLHSLTRDKCGVCGGDDSDCKVVKGTFFNEQGTGGYNVWKIP 790
QY 220 FRGRSRNVKMAIGGKCEKPELLEKACIVGELLOCPRYSWRTSEWKECQVSLLEQ- 278
DB 791 AGSANIDIROGYNMKEDDNYLSLA--ANGEFLN-----GHFQVSLAQOI 837
QY 279 --QD-----PHMHTGP-----YCGGGIQTREV---YCAQSVPAALRA 313
DB 838 AFQDTVLEYSGDALIERINGTPIRSDIYHVLSVGSHPDDISYMTAEPNA----- 892
QY 314 KEYSRVPVEKAL-----CVGPAP-----LPSQ- 334
DB 893 --VIRPISALYLWRVTDWTEDCRACRGQSQKLMCLDMSTHRQSHDRNCONVLKPKQA 950
QY 335 --LCNIPCSSTDICVSSWSAWGLCIHENCHEPQKGFRTQRHVLMESTCPAGCHLVE 392
DB 951 TRMCNIDCSTRWITEDVS-----SCSACSGQKQKRVSCVWEGDRQTPASEHLD 1002
QY 393 --SVPCEDEPMCY-----RWLASGI-CFPDHGKGLGHRILKAVCONDRGDEVSSLC- 442
DB 1003 RNSKPSDIASCIYDCSGRKNYGEWTSCTSGSNGKMR--KSYCVDDSNRRVDESLOG 1060
QY 443 -PVPPPPERKSCIPCRMDCVLSEWT--FWSSCSQSCSNKNSDGKQTRSRITLALAGE-- 497
DB 1061 REQKEATERECNRIPC-----PRWYGHMSECSRSCDG---GVWRHAQCLDAADRET 1110
QY 498 GKKPCPSQALQEHRLCNDHSCMLHWETSPWGPCSEDTLVLTALNATIGNGBEATCGVI 557
DB 1111 HTSREGPAQT-QEH--CNEHAC--TWQFGWSDCS-----AKCGDGV 1148
QY 558 QTRRVFCVSHVQGVMTKRCPPDSTRPETVRPCLPKKDCIVTAFSEWTPCPRMQ----- 613
DB 1149 QVRDANCTDHRSLVPEHRLCKMEKIT-KPCH--RESCPKYKLGWQSCSVSCDGS 1204
QY 614 -----AGNAT-----VKQSYRIRIIIOEAANGQEQP----- 639
DB 1205 SRRVSCVSGNGTEVMSLCTASDRPASHQTCMLG--TCPWRNTDWSACSVSGIGRHR 1263
QY 640 --DTLYEERECEDEVSLC-----PVYRWKPKQKSPCILVPESVWQGTGSS 682
DB 1264 TTECIYREQSV-DASFGDTKMPETSQTCHLLPCTSKPKSHWSPC-----S 1308
QY 683 EACGKGLQTEAVSC-----ISDNRSAEMME-CLKQTNGMPLLVQECTVPCRE 729
DB 1309 VTCGSGIQTNRSVCTRGSGGTIVDEYFCDRNTRPRUKTKCEKDTCDGPRVLQK-----LQA 1364
QY 730 DCTFTAWS--KFTPCSTNCEATKSRRLQTKSRKKEKQD-----SDLYPLVET 777
DB 1365 DVPPIRWATGFWTACSATCG-----NGTORLLKCRDHVRDLDPDEVCHLDEKST 1415
QY 778 ELCPCEFTISQYGNWSDCILPEGREPHRGRLVQADSKECEGELRFRVAVASDKNGRPV 837
DB 1416 RNCRLRDCSYWMAEWEBCPATCG--THVQQRNVTCVSAEDGRTILKDVDC-DVQKRP 1473

```

```
QY 838 DPFSSSGYIQKCVIPCP-FDCKLSDW--SSWSSSSSCGIGVIRSRKWLKEKPYNGG 894
D 1474 SARNCR-----LBPCKGEEHIGSWIGDSKSGASCGGWRRRSVCSSTSSCDET 1524
QY 895 RPKCLDLQKVHEAVPCYSCNO-----YSWVVEHSSCKINNELSLRSLCGGQTOS 946
D 1525 RK-PKX-----FDKNEELCPPLTNSWQISPHWTHC-----SVSCGGGVQR 1564
QY 947 RKICVNTADGEGAVDSNLNQBDEI-----PPTQSC-----SL 981
D 1565 RKIWCEVDLSGR-----KDDIECSSEIKPREQDCMPPCRSHYHNKTSASMTSL 1615
QY 982 MCPNECVMS-----EW-GLASKCPQSCDPHTMORTRHLLRPSLN-----S 1021
D 1616 SSSNSNTSSASASLPLPVWSQTSAMSACAKGRGT-KRRVVECVNPSLNTVTAS 1674
QY 1022 RTCABDSQVPC-----LLNENCFOFQVNLTEWSTQLSENAPQGGVTRTLLSCVSDGX 1077
D 1675 TEC--DQTKPVEEVRCRTKEC--PRWKTITWSSCSVT---CGRGIRREVQCYGRKN 1726
QY 1078 PVMDOCEQHNLEKQFQMSIPCLVECV-VNCLSGM--TATWBCSQTCGHHGRMSRTRFI 1134
D 1727 LVDSSECN-----PKTKLNSVANCFFVACPAYRMNVTPWSCKDECAQKQOTRRVHC 1779
QY 1135 IMPTQGEGRPCFTLTQKCT-----CPVTPC-YSWVLGNWSACKLEGDCGEGVQIRSLSC 1189
D 1780 ISTSGKRAAPRWCELARAPTSIRECDTSNCFYEWVPGDWQTC---SKSGEGVQTRVRC 1836
QY 1190 --MVHSGSISHAAGVEDALCGEFPQDSILKQLCSV-PCPGD---CHLT-----EWS- 1236
D 1837 RKNFNSTIP--IFWLED-----BPAVPEKCELFPPKPNESQTCELNPDSEFKWSF 1888
QY 1237 -EWSTCELTCT-----IDGRSFETV 1254
D 1889 GPWGECSKNCQGIIRRRVKVANDGRRVERV 1920

RESULT 10
QXSV8 PRELIMINARY; PRT; 1637 AA.
AC QXSV8;
AT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE SCO-spondin (fragment).
GN BOS-SPONDIN.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_taxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Gobron S., Creveaux I., Didier R., Meinzel R.;
RT "Characterization of cattle SCO-spondin.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ133488; CAB46239.1; -.
DR HSP; P56882; ICCV.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR001545; Gly_hormoneB.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR008085; TSP 1.
DR InterPro; IPR001007; VNF_C.
DR Pfam; PF000007; Cys_knot; 1.
DR Pfam; PF01826; TIL; 3.
DR Pfam; PF00090; tsp_1; 11.
DR Pfam; PF00093; vwc_1.
DR PRINTS; PR01705; TSP1REPEAT.
DR SMART; SM00041; CT; 1.
```

```
DR SMART; SM00068; GHB; 1.
DR SMART; SM00209; TSP1; 11.
DR SMART; SM00214; VWC; 2.
DR PROSITE; PS01225; CTCK 2; 1.
DR PROSITE; PS00261; GLYC_HORMONE_BETA_1; 2.
DR PROSITE; PS00092; TSP1_11.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 1.
FT NON_TER 1
SQ SEQUENCE 1637 AA; 173347 MW; 4CSBABIDD346C925 CRC64;

Query Match 8.0%; Score 661; DB 6; Length 1637;
Best Local Similarity 22.1%; Pred. No. 4.3e-53;
Matches 353; Conservative 126; Mismatches 516; Indels 604; Gaps 90;

QY 10 RTVANEICEHPALOPPTQACLPDC--VWSEFLPWSNCSKGC-GKKLOHTRAVIA 67
D 269 RTGCHNCSCAHGLKSCSYE-----ACSKAAGGFPWGPWPCSRSCGGLGTRTSRQCVR 323
QY 68 P-PLFGGLQC--PNLTESRACDAPISCPLGEEVTFSLKVG-----PWSKRLPHL 115
D 324 PNPATGGGCHGPH-WDLEYCPSP-ECP-GAAGSTAEPATGLPGWGLWSWSPSCGTCT 380
QY 116 KEINPSGRTVLDNDSNERTVTFKHQSYKAHHKSWAIEIGYQTRQVSTCT--RSDQN 172
D 381 DPAHPAWR-----SRRLCL-----ANCTGGAASQERP 408
QY 173 AMLSCLQDSPLTVQSCIMPKDCETSONSWSPSCSKTCRSGSLLPGRSSRNVMGMAI 232
D 409 CNLPCTE--LPLCPGCEAGNCSWTAWAPWPCSRSCGVG-----QORRLAYHPPG 460
QY 233 GGGKECPPELL-----EKEACIVEGELLQOCP-RYSW-RTSEWKBCQVSLLEQQPHMHT 286
D 461 PGCHWCPCDVLTAQERRFC-----NLRACPVGGSWSPMSWCDRS-----502
QY 287 GPVCGGGIQTREYVCAQSVPAALAAKAEVSEPVYEK---ALCVGPAPLPOLCH-IPCS 342
D 503 ---CGGG-----RSLSRSCSPPPKNGGAPCVGERH-HAKLNTPTCEE 543
QY 343 DCIVSSWSAWGLCIHENCFPGKKGFTQRHVLMESTGPAGHCPH---LVESVPC-E 397
D 544 GC-----PAG-----MEVVSANCRPRCSDLQEGIVCOE 573
QY 398 DPMCYRW-----LASGICFP-DHGKC--GLGHRIL-----KAVCON--DRGEDVSGS 440
D 574 DQACQCGCRCEGSLQDGGCVPLGHCECTDAQGHSWAPSGHQEACNCTCRAGUSCT 633
QY 441 LCPVPPPPPERKSCIEPCRMDCVLSWETWSSCSOSCNKNSDGKQTRSRITILALAGEGK 500
D 634 AQPCPPPAH-----CAWSEWSAWSPCSRSC--GPAGQOSRFRS--STSGSWAP 677
QY 501 PCPPSQALQEHRLCNHSCMLHWETSFWGFCSEDTLVTALNATIGNGEATCGVGIQTR 560
D 678 ECREEQSOS-----QPCQSPC-----694
QY 561 RVFCVKSHVGQVMTKRCPDSTRPTEVRCFL-----PCKKDCIV-----TA 601
D 695 -----PPLCLQGTFRSLGSLWLDQCCQCSCTPFGIICEDACAGLGAWTP 741
QY 602 FSEWTPCPRMCOAGNATVQSKRYRIIOEA-ANGQEC--PDTLYBERECEDVSLCPVY- 657
D 742 WSPSDDCPVSCGGN-----QVRTRVCVASAPRGGSPCLGPDV--QSQC-GLMPCPALP 794
QY 658 ---RWKPKWSPCILVPESVWQGTSGSEACGKGLQTRAVSCISDNNESAEMWELK--- 711
D 795 DTCSWGP--WGPC-----SRSCPGGLASRASC-----PCLLAEA 827
QY 712 ----QTNGMPLLVQECTV-PCREDCTFTAWSKFTPCSTNCEATKSRREQ----- 755
D 828 EPACNSTSPRLDTQACYAGPCLEECVWSSWSWTRCS--CEVLVQQRVHRQHPAPGGAGA 885
QY 756 -----LTGSKRKEKQDSLDLYPLVETELC--PCDEFISQPYGNW----- 793
```



```
b 886 GPPCRLDGHFRPCLTG-----NCSEDCAPPEFQACSGPCTGLCATYLSPLWCQDLP 939
Y 794 ---SDCIILPEGRREPHRLVQADSKEC-----GEG-----LRPVAACDKNGRP 836
b 940 PCQPCYCEGLEAGGC-VPEPCNCHVSGEGAGVTLAPGRDLQGCCECQGE- 997
Y 837 VDPSCSSGYIOEKCVIPCFDFCKLSWWSGSSCGGIVRIRSKWLKEKPYNGRP 896
b 998 ---LQCTSGQ-----CQGLPLSGWSEW-----SPCG-----P 1022
Y 897 CPKLDLKQVHEAVPCYSECNQYVWVHWSCKINNELRLSLRCGGGTQSKIRCVANTAD 956
b 1023 CLPICLLAPASRAA-----LERWQDTRAGLSPTAPTLASEQHRLHLCIDPET 1071
Y 957 GEGGAVDNLC-----NQDEIPETQSCSLMPCNECMSEWGLWKCPCOSCPHTMQRRTR 1012
b 1072 GRPWAGDPDLCTVPLSQORLCPDPGACQDLQ-----WGPFWGAWSPCQVPCGGFRLAWRE 1127
Y 1013 HLIRP-----SINRTEC-----AEDSQVQCLLENCFQFQYNLTWSTCOL 1054
b 1128 AGIPPGGGRGFWAOTESCNGMPCPGSECEAQDTVPTPCANQ-CPRSCVDLMDRVECL- 1185
Y 1055 SENAPCGGQVTRLLSLCVCSGDK-----PVSMDOCEQHNLKSPORMSIPCLVECV-V 1105
b 1186 --QGFCRFG-----CRCPQLVQDGHCVFVSSRC-GLPSFNASWALAPAEVURL 1233
Y 1106 NCQ-----LSGWTATECOTCGHGRMSRTRFIIMPTQGG-RCPC 1146
b 1234 DCRNCTCVNGSLACSHRCPTLGPWSANWSAPCG-GGTTKRHRSC---KEGPGVTPCQ 1289
Y 1147 T-ELTQEKTCPTPCYSWVLGN-WSACKLEGGDCGEGVQIRSLSCMVHSGSISHAAGRVE 1204
b 1290 AQDMEQQDCNLCQPCPCPPGQVLSAC-----AVSC----- 1320
Y 1205 DALCEMFPQDSILKOLSVPC-----PGDCHLITE-----WSEWSTCELTLC 1245
b 1321 PRLCSHLQFGTFCMQEPCQGLGCDPCRGQLLHNGTCVPPAECPCTQLSLPWLGLTLE--- 1377
Y 1246 IDGRSFETVGRQSRRTFIQSFENQDSCPQQLFTRPTCGKCVHYTWKSLANNERT 1305
b 1378 -----EGHRELPGTLLTQNTCHICICOGGAFSCSLTD----- 1409
Y 1306 VMCQRSDGVNVTGGSP-----QARPAIROCIPACRXPFSYCTQGGV-----CGCEK 1353
b 1410 --CQB-----CPEGTWQVAPGELGCEQTCREPNATETQGNCSGRQAPGVCQOR 1458
Y 1354 GYTEMKNSGFLDYCMKV-----PGSEDKXA 1379
b 1459 GHFRSQEGP-----CVPVDLCEWGHGRPHHPGSEWQKA 1492
```

RESULT 11  
D 075851 PRELIMINARY; PRT; 4123 AA.  
C 075851;  
T 01-NOV-1998 (TrEMBLrel. 08, Created)  
T 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
T 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
E WUGSC:H\_DJ0751H13.1 protein (Fragment).  
S WUGSC:H\_DJ0751H13.1  
S Homo sapiens (Human).  
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
M Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
X NCBI\_TaxID=9606;  
I [1]  
I SEQUENCE FROM N.A.  
A Leonard S., Graves T., Stromatt C.;  
T "The sequence of Homo sapiens PAC clone RP4-751H13.";  
I Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
I [2]  
I SEQUENCE FROM N.A.  
A Waterston R.;  
I Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.

```
DR EMBL; ACC04877; AAC36301.1; -.  
DR HSSP; P01130; 1AJJ.  
DR GO; GO:0005507; F.copper ion binding; IEA.  
DR GO; GO:0005489; F.electron transporter activity; IEA.  
DR GO; GO:0004867; F.serine protease inhibitor activity; IEA.  
DR GO; GO:0007155; P.cell adhesion; IEA.  
DR GO; GO:0006118; P.electron transport; IEA.  
DR InterPro; IPR000923; BlueCu.1.  
DR InterPro; IPR001064; Crystallin.  
DR InterPro; IPR006209; EGF like.  
DR InterPro; IPR000421; FAS8 C.  
DR InterPro; IPR002223; Kunitz_BPTI.  
DR InterPro; IPR002172; LDL_receptor_A.  
DR InterPro; IPR002919; TIL_Cysrich.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR008085; TSP 1.  
DR InterPro; IPR001007; VWF_C.  
DR InterPro; IPR001846; VWF_D.  
DR Pfam; PF00754; F5_P8_type_C; 1.  
DR Pfam; PF00057; ldl_recept_a; 1.  
DR Pfam; PF01826; TIL; 5.  
DR Pfam; PF00090; tsp_1; 14.  
DR Pfam; PF00093; vwc; 1.  
DR Pfam; PF00094; vwc; 3.  
DR PRINTS; PRO0261; LDLRECEPTOR.  
DR PRINTS; PRO1705; TSPREPEAT.  
DR SMART; SM00231; FAS8C; 1.  
DR SMART; SM00192; LDLA; 10.  
DR SMART; SM00209; TSP1; 14.  
DR SMART; SM00214; VMC; 1.  
DR SMART; SM00216; VWD; 3.  
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.  
DR PROSITE; PS00196; COPPER_BLUE; 1.  
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.  
DR PROSITE; PS00022; EGF_1; 1.  
DR PROSITE; PS00022; FAS8C_3; 1.  
DR PROSITE; PS01209; LDLRA_1; 9.  
DR PROSITE; PS00068; LDLRA_2; 9.  
DR PROSITE; PS00092; TSP1; 14.  
DR PROSITE; PS0184; VWF_C_2; 1.  
FT NON_TER 1  
SQ SEQUENCE 4123 AA; 434981 MW; 7AAB6F8BDC8012FB CRC64;  
  
Query Match 8.0%; Score 655.5; DB 4; Length 4123;  
Best Local Similarity 21.9%; Pred. No. 4.9e-52;  
Matches 390; Conservative 149; Mismatches 567; Indels 677; Gaps 101;  
  
QY 28 EQALIPCRDVCVSEFLPWSNCKGCKKLOHRTAVIAPLFGQLQCP-NLTSSRAD 86  
DB 2507 EDGCV-----DCVLAPWSVWSSCSRSCGLGLTFQRELLRPPLPG-SCPRDRFRSQSCF 2560  
  
QY 87 APISCFLGEETFTSLKVGPSKCRPLPHLKEINPSGRTVLDFNSDSNERNVTFKHQSYKAH 146  
DB 2561 VQ-ACPV-----AGANWAEANGPCSVSCG----- 2585  
  
QY 147 HHSKSWAIEIGYQTRQVSCSTRSDGQN--AMLSLCLQDSPLTVQSCIMPDCETQSWS-- 202  
DB 2586 -----GHQSRQSRSCVDPPEKNGGAPCPGASQERAPCGLQPCSGGTGKVLGWGHG 2635  
  
QY 203 -----SW-----SPCSKTCS-----GSLLPGRFSRNVK 228  
DB 2636 GSTVGTGRLGLPAPRLTWCPSPTLRAGPCVCEKSVPGAGAPMPTLLFG--SQQQKL 2693  
  
QY 229 HMAIGGG-----KECPLELKE----- 245  
DB 2694 QWALCGSLPSLLCPLGLSLFALPHILPCRCPPGLLLDHTRCLPLSECFCLVGBELKWPVGS 2753  
  
QY 246 -----ACTIVE-GLLQO---CP-RYSWRT-SEWKECVSLLLEQODPHWHTVGPVCGG 292  
DB 2754 FLIGNCSQVCCKGELLQCPGGCPLPCGWSAWSNAPCDRS-----CGS 2797  
  
QY 293 GIOTREYVCAQSVPAALRAK-EVSRP-----VEKALCVGPAPLPSCNIP 339
```

2798 GVRAR--FRSPNPAAWGAPCEGRQELQGGCHTVCGTGIAGISLGAAGVPPSSQFCTLR 2855  
340 C-----SDICV-----SSWANGLCIHENCHEPQKKGGETRQRFVLMESTGPAGHC 387  
2856 THGMPTDHTGTGIEVPGTFTWTSWSSC-SQSLAPAGPGWMSR--LCPSPGDSS-C 2911  
388 P-HLVESVPCEDPM-CYRWLASGICFPD-----HGK-----417  
2912 PGDQTEPCSPPIECTGFCAPGCTPPGLFLHNASCLPRSQPCQLHQLYASGAMARL 2971  
418 -----CGLGHRLKXAVCONDRGEDVSGSLCVPPPPPKSKCEIFCRMDCVLSEWTS 470  
2972 DSCNNCTCVSG-----KNACTSER-----CPVA-----CGSPWTLWS 3004  
471 SCSSCKNSKNSDGRQTRSRILA-LAGEGKPCPPSOALQEHRLCHNDHSC-----M 520  
3005 LCSCSC-----NVGLRRFRAGTAGTAPPAFGAEC--QGPTEAEFCSLRPCPGVPWCMCR 3058  
521 QLHWETSPWGP--CSEDTLVLTALNATI-----545  
3059 DKQLDCAQGPASCALASAPRTNQTCHPGCHCPGSMGLLLVSPRHPGLGASVQPPVAL 3118  
546 -----GW-----NGEATCGVGIGIOTRRVFC-----VKSHV 569  
3119 PGAIGTGVGAGGWPWNSHSCSGGGLASRTFACDPPQGLGDYCEGPRAQGEV 3178  
570 GQVMTKRCP--DSTRPETVR--PCFLPKKDCIVTAFSEWTPCPRMCOAGNATVQKSYR 625  
3179 QCALP--CPVTNCTAIGABEYSPGCPPRSCDDLHVCVWR--QPGCYCPGPGVLSNGA 3235  
626 IIOEAAANGOECPDTLYEER-----ECEDVSLCVFYRWKQP 662  
3236 ICVPQ--GHCSLUDLTGQKHHPGARLAPDGNCHCTLEGRNCTDLP-CPDC-GGGQ 3290  
663 KWSPC-----ILVPESVWQ-GITGSSEACG-----KGLQTRAVSICSDNRSR 704  
3291 SLHFCGQPCPRSCDLSGSPGVCQPGVCGOPTCGPLGQLSQDGLCVPAHC-----3342  
705 EMMECLQKTMG-PLLYQECTV-----PCRE--DC-----TETANSKFTPCSTNC--EATK 750  
3343 ---RCQYQPGAMAFSPVSTCVAGILOQCEVPDPCDPGVWSSWGFWEDCSVCGGGEQLR 3399  
751 SRR-----RLQTKSRKKEKQSDIYPLVET--ELCPD-EPISQFYGNMSD 795  
3400 SRRCARPPCPGAPASQRTCSQVCREAGCPAGRLYRECQPGEGCFPSCAHTVQVGCFS 3459  
796 ---CILPEGRREPHRLRVO-----ADSKCEGELR 823  
3460 GCEGCHCEGTFO-HRLACYQECPCVLTAWLLOELGATIGDPGQPLGPGDELDSQTLR 3518  
824 FRVACSDKNGR--PVDPSFCSSGY-----IQBCKVIPCPF--858  
3519 TSCGNCSAHGKLSCLDDCEADGGFGPWPSPGCSRSOGLGTRTRESQCVLTWPTLS 3578  
859 -----DCKLSDWSWSGSSSSCGIGVIRISKMKEXYPNGRCPKLDLUNQ 905  
3579 BLPVCPGPGGAGNCSWTSWAFEPWPCRSRSGVQGRRLRAYR--PPGPGHWCNPILATYQ 3637  
906 VH-----EAVPCVSECNQYSWVVEHWSCKINNEL-RSLRCGGGTQSR 947  
3638 ERRFCNLACBACGAGPAGMEVTTANCRPCRSDLQEGIVQDDQVQCKGCRCKPSLEQ 3697  
948 KIRCV-----NTADGEG-----GAVDSNLCNDEIPPETQSCSLM-C--PNECVMSWGL 994  
3698 DGGCVPIGHCDCTAQGHSAWAPGSGHQACNCSQAGQLSCTAQPCCPPPTCAWSHSA 3757  
995 WSKCPQSDPHTMOTRTHLLRPSINR--TCAEDSQVQPCLL-----NE 1037  
3758 WSPCHSGCPGQGRFR--CGPLASGSGC-----PCMAKADPTCNSTFLHLDTO 3808  
1038 NCF-----QFYNL--TEWSTCOLSENAPCQGG--VTRLLSCVCSDG--KPVSMDCQEQ 1086  
3809 GCYSGPCPDSQWSLWGPWSPQV-----PCSGGFLRWREAEALCGGSCREPWADR---3861

1087 HNLKPPORMSIP--CLVECVV-----NCOLSG-WTAWTECSOTCGHGRMGRTRFIIMPTQ 1139  
3862 ---XLRRALPSTCVNESILVCPHQECPLVGPNSWSSCSAPCG-GGTMEHRHC-----EG 3913  
1140 GEG-RPCELT-GEKTCFVTPCVSWVLGN-WSACKLEGGDCGEGVQIRLSQM-----VHS 1193  
3914 GPGVAPCAQTEQRCNQLQPCPECPFGVLSAC-----ATSCPLCWLQOP 3961  
1194 GSISHAAGRVEDALCGMPFODSILKQLCSVPQDCHLTSEWSEWSTCELTICIDGRSFET 1253  
3962 G-----ALCVBPCQPG--CG--CPGGQHSILPWGLTILE-----3992  
1254 VGRSRSRTFIQSFENQDSCQVLETRFCTGKCYHYTWKASLMNNERTVWCQRSDG 1313  
3993 -----EOAQELPPGTVLTRNCTRCVCHGAFSCSL-----VDCQGE--4028  
1314 VNVTCGSCSPARPAARQIPAC-----RKFFSYCTOGGVC 1351  
4029 IVPGETWQVAPGELGLCEQTCLMNAATKQSCNSARASGC 4071  
  
RESULT 12  
Q8MYA8 PRELIMINARY; PRT; 1461 AA.  
AC Q8MYA8, Q17591, Q22300;  
DT 01-OCT-2002 (TRENBLrel. 22, Created)  
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE C. elegans ADT-1 protein (Corresponding sequence C02B4.1).  
GN ADT-1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN (1)  
RN SEQUENCE FROM N.A.  
RA Kuno K., Baba C., Asaka A., Hosono R.;  
RT "Analyses of C. elegans ADAMTS family gene, adt-1";  
RN Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RN SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C. elegans: A platform for  
investigating biology";  
RL Science 282:2012-2018 (1998).  
RN (3)  
RN SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA McMurray A.A.;  
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB066246; EAC05514.1;  
DR EMBL; Z50004; CAA90293.2;  
DR EMBL; Z50006; CAA90293.2; JOINED.  
DR EMBL; Z50004; CAA90302.2; JOINED.  
DR EMBL; Z50006; CAA90302.2;  
DR PIR; T18856; T18856.  
DR GO; GO:0004222; P-metalloendopeptidase activity; IEA.  
DR GO; GO:0008270; P-zinc ion binding; IEA.  
DR GO; GO:0008508; P-proteolysis and peptidolysis; IEA.  
DR InterPro; IPR001590; Peptidase\_M12B.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR InterPro; IPR008084; TSP1.  
DR InterPro; IPR008085; TSP1.  
DR Pfam; PF01421; Repolysin; 1.  
DR Pfam; PF00090; tsp\_1; 12.  
DR PRINTS; PR01765; TSP1REFAT.  
DR SMART; SM00209; TSP1; 13.  
DR PROSITE; PS00215; ADAM\_MEPPO; 1.  
DR PROSITE; PS00092; TSP1; 12.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.

Q SEQUENCE 1461 AA; 162616 MW; 3CFDC1C07C1F493A CRC64;  
Query Match 7.1%; Score 587.5; DB 5; Length 1461;  
Best Local Similarity 23.4%; Pred. No. 3.7e-46;  
Matches 315; Conservative 155; Mismatches 460; Indels 419; Gaps 86;  
Y 12 VYANEI-----CEHFALQPTQACILPCPRCIVSEF-----LPWNSCS----- 51  
b 385 VLAHEHGHNGMVDHGVQNCNKG-----CLMSAVGAGKTTWSDCSVREFNAFLQ 437  
Y 52 ---KGGCKLQHRTRAVIAPFLFGLQCP--NLTSRACDAIPSCPLGEBEETFLSKVGP 106  
b 438 LDESSEGNCLRASPCILSTNHLSDLRLPQGFADQCC-----SYFWR 482  
Y 107 WSKRLPLHKEINPSGRTVLDFNSDSNERVTFKOSYKAHHHSKWAIEIGVQTRQVST 166  
b 483 DYKVEIPNGKAMDDICRIWCSNGST--ISTAHPALEG-----SWC----- 522  
Y 167 RSDGQNAML--SLCLQDSPLTVQSCIMPKDCETQWSSWSPCSKTCR-----SGSLLP 218  
b 523 ---GANKWCHKQCTHTWTFGLTP-----VFID-----GEWSEWGAEGKCPICQCAVSGSITV 572  
Y 219 GFRSRNRKMAIGAIGGKECPLELKEAC-----IVEGELLOQCPRYSWRTSEWKECQVSL 274  
b 573 QGHRD-CVNPAPNNGKTC-----EGANIRGIVCGATSSNCLGFTREFGNKICS-SI 624  
Y 275 LLEQDPHWHVTGPVCGGIGIOTREYVCAQSVPAALRAKEVRSRVERKALCVGPAFLPSQ 334  
b 625 KYDPHPDQOLYGEFGEHTQPCRVMC--HLTGSLEIRNG----- 663  
Y 335 LCNIPGPOSTCIVSSWSAMGLCIHENCHEPQGGKPRTRQHRVLMESTGPAGCHPLVESV 394  
b 664 ---QFPDGTFCGDAVCVGOCLALSC-----DNKALVFO---PEDCPRI----- 702  
Y 395 PCEDEWCYRW--LABEGICFPDHGKCGLCHR--ILKAVQONDRG--EDVSGSLCPVPPPPER 450  
b 703 ---EGRSVHOWEWSWSEC-----SVSGGLGREGVREKCSGKQCVS-----EES 748  
Y 451 KSCETPCRMDC-VLSEWTEWSSCSOSCKNKGDKQTSRTILALAGGKPCPSQA-- 507  
b 749 RCEGVLR-DCEFEWKEWGSCEKAL-----GVQKPR-----PLTDQCS 792  
Y 508 --LOEHLRNDHSCMLHWETSPWGPCSDETLVTALNATIGNWGATCGVGIOTRRVFCV 565  
b 793 KHLQERPCDNEGCW-TNWD--EWSSCSQ-----SCGGRRYR----- 827  
Y 566 KSHVGQVMTKCP-DSTRTPETVRPCLPCKDCIVTAFSEWTPCPRMQAGNATVQSBRY 624  
b 828 ---IRKCLDDKCDGDDLEKESCN-----TQKISQSGMDLPLCSVSCGIG----- 869  
Y 625 RIIOEAAANGQCEPDTLYEERECEDEVSLCPVYRWKPKQKSPCILVPESVMQGITGSEA 684  
b 870 -FQIRELCDGELCATANKQARTCQ-QQCP-----SAFSLSVWSEWGTTCSAT 918  
Y 685 CKGKLOT-----RAVSCISDDNRSAEMCECLKQTMGLLVQECTVPCREDCFTTAWSKF 739  
b 919 CGEGLQSRSCRRSGCTEDD--ASQTRCV--NG-----PCEH--SVLTWSEW 961  
Y 740 TPCSTNCEATKRRRLQTKSKRKEKQCDSDLYPVVETELPCDEDFISQPYGNWSDCILP 799  
b 962 TTCTET-CSSFDGRKR-IAKCDGTENCQDK-----IDBETCDI-----ACL-- 1000  
Y 800 EGRREPHRLGRVQADSKCEGEGELFRVAVACSKN---GRPVDPSFCSSSGVIOEKCVIPC 856  
b 1001 ---REKHSFGPISP-----RRPLITSNLRKAFGRPLLPISHS-----EK----- 1040  
Y 857 PFDCKLSDWSSGSCSSCGIGVIRSKWLEKPKYNGGRPCPKLDKNQVHEAVPCYSEC 916  
b 1041 -----WSEWGPCSVTCSGRRVATRCQE-----ASCPEHQIQT-----EC 1077  
Y 917 NOYS-----WVVEHWSSCKINNELSLRCG-GGTQBRKIRCV-NTADGGGAGVDSNLCNQD 970  
b 1078 NLNSCLELFIWSDWSSC-----SKSCQDGIQTRQKLCLENNAECSYA-ESRCK-- 1127

QY 971 EIPPETQSCSLMCPNECV-----MSEWGLWSKPOSCDPT-MQRRTRHLLRPSLN 1020  
b 1128 DLP-----SCSSISSGRITSENGFDPARWSEWSWSAC--SCFSLTSTRRFFCQVDPVTVQ 1181  
Y 1021 SRCAEDSVQVQPCLLNENCFQFO--YNL-TENSTCOLSENAPOCGQVTRLLSCVSDGK 1077  
b 1182 GFCAGAILQIPCAPG--SCSPSAGGSHLSEWSSCS-KDCGDTGHQIRNM-----CSEPI 1235  
Y 1078 PVSMD-QCEQHNLEKFORMSIPCLVECV-----VNCQLSGWTAWTECSOTCGHGRMSRT 1131  
b 1236 PSNRGAYCSYSD--QR---PCVMNVCSDEKVDGWDTDWASECTDYCRNGHR-SRT 1289  
Y 1132 RFIMPQGGREGPCTELTOEKTCPVTPCY-----SWVLGNWSACKLGGDC 1178  
b 1290 RFCANPKPSGGAQCTGSDFE---LNPFCDPARCHLRDGGWSTW--SDWTPC---SASC 1340  
Y 1179 GEGVQIRLSLSCMVHSGSISHAAGRVEDALCGEMPFODSILKQLCSVPC---PGDCHLTEW 1235  
b 1341 GFGVQTRDRSC---SSPEPKGQ---SCSGLAHQTS-----LCDLPACDHESDGEWSAW 1388  
Y 1236 SEWSTCELTICIDGRSFETVGRQSRRTFI 1264  
b 1389 NEWSGCMGNC-----GIGTRTRVRACV 1410  
RESULT 13  
ID O43376 PRELIMINARY; PRT; 182 AA.  
AC O43376;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE Hypothetical protein D1110N13.1 in chromosome 7 (Fragment).  
GN D1110N13.1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Andrews S., Holmes A.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004141; AAC02753.1;  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00030; TSP\_1; 1.  
DR SMART; SM00209; TSP1; 1.  
DR PROSITE; PSS0092; TSP1; 1.  
KW Hypothetical protein.  
FT NON TER 1  
FT NON TER 182 182  
SQ SEQUENCE 182 AA; 20189 MW; B898DF494381AB71 CRC64;  
Query Match 6.7%; Score 555.5; DB 4; Length 182;  
Best Local Similarity 52.8%; Pred. No. 2.5e-44;  
Matches 95; Conservative 29; Mismatches 56; Indels 1; Gaps 1;  
Y 577 CPDSTRPETVRPCLPCKKDCIVTAFSEWTPCPRMQAGNATV-KQSRRIITQEAANG 635  
b 1 CPBSLRPETVRPCLLCKKDCIVTAFSEWTPCPRMQAGNATV-KQSRRIITQEAANG 60  
Y 636 QCEPDTLYEERECEDEVSLCPVYRWKPKQKSPCILVPESVMQGITGSEHACGKLOTRAVS 695  
b 61 RDCTDPLYEKACEAPQACQSYRWKTHKWRRCCLQVPSVQDQSPGAEQCGGCGRQARIT 120  
Y 696 CISEDNRASMMELQTMGPLLVQECTVPCREDCFTTAWSKFTPCSTNCEATKRRRQ 755  
b 121 CRKQDGGAGIHECLQVAGVPVPAITQACQIPQDDCQLTSWSEKSSCGDCCGAVTRKRT 180  
Y 756 L 756  
b 181 L 181

RESULT 14

Q8T3A0 PRELIMINARY; PRT: 1089 AA.  
 ID Q8T3A0  
 AC Q8T3A0; (Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE Putative coagulation serine protease.  
 GN SP4  
 OS Ciona intestinalis.  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
 OC Phlebobranchia; Clonidae; Ciona.  
 OX NCBI\_TaxID=7719;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=Hepatopancreas;  
 RA Hammond J.A., Nakao M., Yano T., Kemp G.D., Smith V.J.;  
 RT "Complete cDNA sequence of a serine protease from Ciona  
 RT intestinalis";  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ431686; CAD24309.1; -;  
 DR HSSP; P00761; IAN1.  
 DR GO; GO:0016020; C-membrane; IEA.  
 DR GO; GO:000509; F:calcium ion binding; IEA.  
 DR GO; GO:0004263; F:chymotrypsin activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0005044; F:scavenger receptor activity; IEA.  
 DR GO; GO:0004295; F:trypsin activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR009003; Cys\_ser\_trypsin.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF-like.  
 DR InterPro; IPR006210; IEGF.  
 DR InterPro; IPR001254; Peptidase\_S1.  
 DR InterPro; IPR001314; Peptidase\_S1A.  
 DR InterPro; IPR003582; SHK1.  
 DR InterPro; IPR001190; Stcr\_receptor.  
 DR InterPro; IPR008884; TSP1.  
 DR InterPro; IPR008085; TSP\_1.  
 DR Pfam; PF01549; SHK1; 3.  
 DR Pfam; PF00089; trypsin; 1.  
 DR Pfam; PF00090; tsp\_1; 11.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PRINTS; PR01705; TSP1REPEAT.  
 DR SMART; SM00181; EGF; 1.  
 DR SMART; SM00254; SHK1; 3.  
 DR SMART; SM00020; Tryp\_Spc; 1.  
 DR SMART; SM00209; TSP1; 11.  
 DR PROSITE; PS01186; EGF\_CA; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS0287; SRCR\_2; 1.  
 DR PROSITE; PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE; PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
 DR PROSITE; PS00092; TSP1; 11.  
 DR EGF-like domain; Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 1089 AA; 113693 MW; 2742DE15E22F6C0A CRC64;

Query March 5.8%; Score 474; DB 5; Length 1089;  
 Best Local Similarity 20.7%; Pred. No. 1.6e-35;  
 Matches 275; Conservative 117; Mismatches 425; Indels 510; Gaps 73;

Qy 168 SDGQAMLSLCLQDSFFLTVQSCI-----MPKDCETQSWSPSCSKTCRSGILLPG 219  
 Db 21 SQGGSSVVCSTSL-----GCVDCFSWCQANAASCTSSPALMGSYCKXTCN----- 64  
 Qy 220 FRSRNRVKNVAIGGKECPLELKE---ACIVTEILLQOCPRYSWTSWKE-----CQV 272  
 Db 65 -LCASNGAACIAKGNHRCIETTSSEPVCACF-EGFRLEANGRTCTVDIDEAENSTLCS 122  
 Qy 273 SLLEQDPHHVHTGPVCGGGIQTREVV-----CAQSVPAAL-----RAKEV 316

Db 123 PNLNPNNTLGHVYCTACGSPNRYHYVERNECKMGSGACGKSTNGRIVGKRGRIA 182  
 Qy 317 SRP-----VZKALCVGPA-----PLPSQLC-----NIPSTDIV----- 346  
 Db 183 RWPWMAIYVIGRNLCGGTLLSSGWLTAACFASTINNNPSTINVLGVVDTDSGNIHE 242  
 Qy 347 SWSAWGLCIHENCHEPOGKGFTRQRHVLMESTGPAGHC-PLHVESVPCEDPMC----- 401  
 Db 243 QSFVSVTRLIHHPNPNNDLALLQDLHDALIDAAFKVPCVLPNGEE--PPEGKCMWATG 300  
 Qy 402 YWLASEGIC-----FPDHGKC-----GLGHR-----LKAVCONDR 433  
 Db 301 YGTIAFGVAAKSLQEVNLPADLAHCERIYANTNRVNTMLCAGYITGCKDTCCGDS 360  
 Qy 434 GBDVSGSLCPVPPPPPKSC-----IPCRM----- 459  
 Db 361 GGPLVCQRC-----KNCDWYLAGITTSFGRGCARPGFFGVYTKVSFFEQWISSYTSIA 412  
 Qy 460 ----DCVLSEWTEWSS-----CSQCSNKNKSDGKQTRSRITLALAGEGKP-CPPSQALQEH 511  
 Db 413 INPGQCVKPSWTTWGSWTPCA-SCSGSS-----SRIFCANGSPDGPDCGLQ--BEF 462  
 Qy 512 RLCNDHSCWQLHW-ETSPWGPCSEDTLVTALNATIGWGEATCGVIGIOTRRVFCVKSHV- 569  
 Db 463 ROCST-VCTQPTWAEYGDWGS-----VTCGDGSRSRICRNGNIG 504  
 Qy 570 -----QGVMTK-----RCPD----- 579  
 Db 505 DPGCTGGETATEACTTGVRCPTWSAMSGYGVCSVTCGGGTQESTRTCNHNGQAGVTC 564  
 Qy 580 -STRPETVRPCFLPCKKDCIVTAFSEWTPCPMCOAGNATVKQSRRIIIQEAANGQEC 638  
 Db 565 RDRSQACNPQTCFAP---TWAAYGAMSDCTRCGGERT-----RVATCL-NGAIGSGC 616  
 Qy 639 PDLTYERECEDEVSLCPVY-RWPKQ-KWSPCILVPESVMQGITGSSBACGKLOTRAVSC 696  
 Db 617 PAAGVSQTESCNISQCANPTWSAYGWSGC-----SVTCASGTRSRSC 662  
 Qy 697 ISDD-NRSEAEMWELKQTNMPLLVQECT--VPCREDCTFTAKSKTTPCSTNC-----E 747  
 Db 663 VGNIGNVGCE-----SGQTASEACTTGVQCP---TWSAMSVYGVCSVTCGGGTQ 711  
 Qy 748 ATKSRRRQ-----LTGSKRKEKQSDSLYPLVETELCPCEDEFISQPYGNWSCILPEG 801  
 Db 712 STRTCNNHGVGVTCGDRDTRSQAC-----NPOACPS-----WSGYGWSGC----- 753  
 Qy 802 RREPRLRVQADSKCEGEGRLFRVAVACDKNGRPVDPSPFCSSSGYICEK---CVIPCF 858  
 Db 754 -----SETCGDGTKTRTCN--NGQ-IGDNGCSPAAATDSMACSVRNCP- 796  
 Qy 859 DCKLSDWSWSGSCSSCGIGVIRSKML-----KEKPYNGRCPCKLDL 902  
 Db 797 --QWSSWSGWSGCSLTGCGGTRTAVRCNTFGATCASCAGATSKSEPCNLG----- 846  
 Qy 903 KNOVHEAVPCISECNQYSHVVEHSSKINNELSLRCGGGTQSRKRCVNTADGEGAV 962  
 Db 847 -----ACPFVS-----AW--SGWSTC-----SAGCGGGQTRTRTCSFPGNCDPAF 886  
 Qy 963 DSNLNCQDEIPEETOSCSLMCPNECMSEWGLW--SKPCOSCDPHTMQRTRHLLRPSL 1019  
 Db 887 GTAL-----SGSQACNTDAGIGWVNSGTCSAACGPTIQ----- 925  
 Qy 1020 NSRTCAEDSQVQPCLLNENCFCQYNLTWSTCOLSENAPCGQGVTRLLSCVCSDDGKPV 1079  
 Db 926 -TRECIGGTAGP-----NC-----VGSTQQTAACN----- 950  
 Qy 1080 SMDQCEQHNLEKQPMRMSIPCLVECVNQLSGWTAWTECSOTCGHGRMSRTRFIIMPTQ 1139  
 Db 951 -----VAACTGEWNTACTVTCG-AGTQTRSR-----TCS 981  
 Qy 1140 GREGPCP---TELQTEKTCPVTPCYSVWLGNSACKLEGDCGEGVQIR-SLSCMWSHGS 1195  
 Db 982 GEAGRCPGQSAATESQACAASTCASTV-----NDCSNDIDLAPAITCREVA-- 1028

2Y	1196	ISHAAGRVE-----DALC-----GEMP-----FQDSILKQLCSVPCGDCHL-TEWSEW 1230
2b	1029	---VAGVCEQYKDYMDINCIRSCCVFGRNFCSMYRDSLQ-----CPSYRHLCNTLVE 1079
2Y	1239	STCELC 1245
2b	1080	PLCKYTC 1086
RESULT 15		
2Y	2966P9	PRELIMINARY; PRT; 1121 AA.
ID	Q966P9	
AC	Q966P9	
DT	01-DEC-2001 (TtEMBLrel. 19, Created)	
DT	01-DEC-2001 (TtEMBLrel. 19, Last sequence update)	
DT	01-MAR-2003 (TtEMBLrel. 23, Last annotation update)	
DE	Hypothetical protein C36B7.5a.	
EN	C36B7.5.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloderinae; Caenorhabditis.	
NCBI_TaxID=6239;		
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Bristol N2;	
RX	MEDLINE=99089613; PubMed=9851916;	
RA	None;	
RT	"Genome sequence of the nematode C. elegans: a platform for	
RT	investigating biology. The C. elegans Sequencing Consortium.,";	
RL	Science 282:2012-2018(1998).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Bristol N2;	
RA	Minx P., Minx M.;	
RT	"The sequence of C. elegans cosmid C36B7.,";	
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Bristol N2;	
RA	Waterston R.;	
RT	"Direct Submission.,";	
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AC006615; AAK68231.1; -	
CR	WormPep; C36B7.5a; CE25806.	
CR	InterPro; IPR003014; PAN.	
CR	InterPro; IPR000884; TSP1.	
CR	Pfam; PF00024; PAN; 1.	
CR	Pfam; PF00090; tsp.1; 9.	
CR	SMART; SM00209; TSP1; 11.	
CR	PROSITE; PS50092; TSP1; 10.	
3Q	SEQUENCE 1121 AA; 126159 MW; 151B03F47E5BF1BC CRC64;	
Query Match 5.7%; Score 470; DB 5; Length 1121;		
Best Local Similarity 23.8%; Pred. No. 3.9e-35;		
Matches 255; Conservative 88; Mismatches 371; Indels 394; Gaps 62;		
2Y	357	HENCHE-POQKGFRTRQRIVL-----WESTGPAGHCPLHVESVPCEDPMCYRWLASGI 410
2b	197	HQSSHERDRGDDTDROKDRYLXTIDYTKSERPIHPTIHRGQSSQQSASPVASGNP 256
2Y	411	CFPDHGKCGLGHRIILKAVCONDRGEDVSGSLCPVPPPPPKSKCEIPC-RMDCV----- 462
2b	257	CLPNACARTLYQMDTAPCPARGD-----PCAPKE-----PCMNDDCYHVAPPGQ 302
2Y	463	-----LSEWTEWSSCSQSCSNKNSDGKQTRSRITILALAGEGKPCPPSQALQEHRLCNDS 518
2b	303	AAPPVVENTEWSBESASC-----GISLRTRQCL-----GGITCIGPSTIP-----CQVPE 347
2Y	519	CMQLHWET-SPWGPCSDETLVTALNATIGNWGATCGVG-IQTRRFVFCVSHVGQVWTKR 576
2b	348	CAV--WTLSWSBELCT-----ATCGVGIEIQRNV-----COTGR 379

QY	577	CPDSTPBTVRPC--FLPCKKDCIVTAFSEWTPCPNRCQAGNATVQSRVRIIIQBAANG	634
Db	380	NCDDGPSVE-VEACKALLPCP---TWSSWMSWTCGNSC--GKSIERRS-----ICQN	426
QY	635	QGCSPDTLYEERCECDVSLCPVY-RWKPKMKSPCILVPBSVMQGITGSSBACGKLQTRA	693
Db	427	GLLCPGPAITERTC-DKRPCPHSNW--QWKEC-----SKECCAGHTYRN	469
QY	694	VSCI-----SDNRSAEMECILKQTNMGMLLVQBCTVPCREDCTFTANSKFTPCSTNC--EA	748
Db	470	REGIDGVSCEGASEEKILCNQO-----PCPEWSQWTAM---TVCDERCGEES	513
QY	749	TKRRRQLTGKSRKKBKCDSDILYVETELC--PCDEFISQPY-----GNMSDCI	797
Db	514	IRLNR-----KQNAE-----NNNACDGPADQMSCDPYRDCPKWEEGWENADC-	557
QY	798	LPGRREPHRGLVQADSKCEGEGLEFRAVACSDKNRVPDPSPFCSSGYIQBKC-VIPC	856
Db	558	-----STTCGGQTKRLRKCDSGNE-----CSGPBEMRFCQIASC	593
QY	857	PPDCKLSDMSSWGSSCSGIGV--RIR-----882	
Db	594	PY---WGDWTPMSGCVSCGGVCERTRKCIITDFLOLPTLEBLRDSLEKHEAKEALI	650
QY	883	-----SKW-----LKERPYNGRPPCKLDLKNQVHEARVPVYSECNQYSVVY	923
Db	651	ARAKTISKYRRTNETRLSPRPLSPPEQLGG--TCDGPEMETKYCDAGC---CVNWRWT-	705
QY	924	EHWSSC-KINNELRSUR-----CGGQTOS-----RKIRCUNTAD-BEGGAVDSNLGN	968
Db	706	-EWSPCIGCGNEGISKRNRVCSMEGOSPAPPIISFQVRSHRTVYVTSLGGPIGPNILPG	764
QY	969	QDSIPP-----ETQSCSLMCPNE-----CVMSEWGL	994
Db	765	MSIPIVPEIHRGKQVLFGIQATPKCH--CPGDTFTQTRPCLEQNAQCDPVGKCEWSEWGE	822
QY	995	WSKPOSCDPHTWQTRTHLLR-----PSLNSRTC-----ABDSQVQPCLLNENCFQFQ	1043
Db	823	WCGC-MRCRPFGEVR-RPCDRSPIRGGPMRPDSTCDGNGDDSDSQRSCPMEQTCYVNG	880
QY	1044	YNLTWSTQCLSENAPCGGVVTRLLSCVCS-----DGKEPVSMDQCEQHNLEKQPSMIP	1098
Db	881	IGTRFNNAASQDARERONFRGRTGCFYNTAAPQDRTSYNSNRRPSPHSRSDETVPID	940
QY	1099	CLVECVNQLSGWTAWTECSQTCGHGRMSRTRFII-----MPTQGEGRPC	1145
Db	941	SVIGRVQVCHWSKWSRCHD---NSTRERKRFCVGEKSELVSNCECLGKPEHEEPC	996
QY	1146	PT----ELTQKTCPTPYCVSWVLGNWSACKLEGGDCGEGVQIRLSLSCVHSGSIASHAG	1201
Db	997	NATGFIEQNEET-----DREIENKLDKLEEDNAGGDVELHKHKEBLKS-KVESATG	1048
QY	1202	RVEDALCGLMFPQDSLKQLCSVPCGDCHLTWSEWSWSTCELTCDIGRFPETVGRQSRSR	1261
Db	1049	TWO-----CDWTWBSQWSVCTASCGEGR-----RWRRR	1077
QY	1262	TFIIQSFENQDSCPQOVLETRPCTGKCK	1289
Db	1078	-----C-----PCGESKC	1085

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 24, 2004, 01:00:01 ; Search time 53.7866 seconds  
(without alignments)  
1406.151 Million cell updates/sec

itle: US-10-022-710-2

erfect score: 8241

equence: 1 MYRCIQKLNRTVANEICR.....QSTPQPKPLTAYDGDLM 1465

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

earched: 389414 seqs, 51625971 residues

otal number of hits satisfying chosen parameters: 389414

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Issued Patents AA.\*

- 1: /cgm2\_6/ptodata/2/iaa/5A\_COMB.pap.\*
- 2: /cgm2\_6/ptodata/2/iaa/5B\_COMB.pap.\*
- 3: /cgm2\_6/ptodata/2/iaa/6A\_COMB.pap.\*
- 4: /cgm2\_6/ptodata/2/iaa/6B\_COMB.pap.\*
- 5: /cgm2\_6/ptodata/2/iaa/PTCTUS\_COMB.pap.\*
- 6: /cgm2\_6/ptodata/2/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	ID	Description
1	694	8.4	US-09-800-729-155	Sequence 155, Appl
2	657	8.0	US-09-369-364A-13	Sequence 13, Appl
3	477.5	5.8	US-09-800-729-89	Sequence 89, Appl
4	440.5	5.3	US-09-132-769-3	Sequence 3, Appli
5	434.5	5.3	US-09-132-769-1	Sequence 1, Appli
6	434.5	5.3	US-09-640-173-186	Sequence 186, App
7	434.5	5.3	US-09-713-550-186	Sequence 186, App
8	429	5.2	US-09-800-729-217	Sequence 217, App
9	428.5	5.2	US-07-862-021B-10	Sequence 10, Appl
10	428.5	5.2	US-08-313-288B-10	Sequence 10, Appl
11	428.5	5.2	US-09-132-769-5	Sequence 5, Appli
12	428.5	5.2	PCT-US93-03164-10	Sequence 10, Appl
13	419	5.1	US-07-862-021B-12	Sequence 12, Appl
14	419	5.1	US-08-313-288B-12	Sequence 12, Appl
15	419	5.1	PCT-US93-03164-12	Sequence 12, Appl
16	383.5	4.7	US-07-862-021B-14	Sequence 14, Appl
17	383.5	4.7	PCT-US93-03164-14	Sequence 14, Appl
18	340.5	4.1	US-08-918-914-4	Sequence 4, Appli
19	330	4.0	US-09-369-364A-17	Sequence 17, Appl
20	329	4.0	US-08-985-526-3	Sequence 3, Appli
21	311.5	3.8	US-08-313-288B-15	Sequence 15, Appl
22	310.5	3.8	US-09-230-682-2	Sequence 2, Appli
23	306	3.7	US-09-930-872-4	Sequence 4, Appli
24	300	3.6	US-08-185-432-19	Sequence 19, Appl
25	300	3.6	US-08-899-232-4	Sequence 4, Appli
26	286	3.5	US-08-083-590A-20	Sequence 20, Appl
27	286	3.5	US-08-532-384-20	Sequence 20, Appl

28	284	3.4	450	4	US-09-369-364A-19	Sequence 19, Appl
29	284	3.4	2556	1	US-08-185-432-17	Sequence 17, Appl
30	284	3.4	2556	1	US-08-899-232-2	Sequence 2, Appli
31	269	3.3	1170	1	US-08-313-288B-20	Sequence 20, Appl
32	264	3.2	2523	1	US-08-185-432-18	Sequence 18, Appl
33	264	3.2	2523	4	US-08-899-232-3	Sequence 3, Appli
34	261	3.2	2471	1	US-08-185-432-16	Sequence 16, Appl
35	261	3.2	2471	1	US-08-083-590A-19	Sequence 19, Appl
36	261	3.2	2471	3	US-08-532-384-19	Sequence 19, Appl
37	261	3.2	2471	4	US-08-899-232-1	Sequence 1, Appli
38	251	3.0	1964	4	US-09-467-997-1	Sequence 1, Appli
39	250	3.0	525	4	US-09-369-364A-21	Sequence 21, Appl
40	240.5	2.9	1205	4	US-09-491-522-11	Sequence 11, Appl
41	239.5	2.9	1251	5	PCT-US95-02251-3	Sequence 3, Appli
42	239.5	2.9	1252	1	US-08-199-780-3	Sequence 3, Appli
43	239.5	2.9	1252	2	US-08-316-650-3	Sequence 3, Appli
44	239	2.9	1211	4	US-09-491-522-5	Sequence 5, Appli
45	232	2.8	1833	3	US-08-479-722B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-09-800-729-155  
; Sequence 155, Application US/09800729  
; Patent No. 6605592  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: PZ044PI  
; CURRENT APPLICATION NUMBER: US/09/800,729  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: PCT/US00/26013  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: 60/155,709  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 155  
; LENGTH: 2165  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-800-729-155

Query Match	8.4%;	Score 694;	DB 4;	Length 2165;
Best Local Similarity	22.9%;	Pred. No. 4.1e-47;		
Matches	351;	Conservative 133;	Mismatches 556;	Indels 492; Gaps 84;
Qy	18	CEHFALQPTTEQACLI PCPR-----DCVSEFLPWSNCSKGCKLQHRTRAVIAP-PLF	71	
Db	586	CHH-----GACVRLAPESLT KIDGWDNRSGECSTCGGVQKGLRDCDSPKPRN	637	
Qy	72	GGLOCNLTES-RACDAPISCLGBEEYTFSLKVGPKWCKRLPHLKEINPSTRTVLDNFS	130	
Db	638	GGKYCVGQERYSCNTQ-ECFWDTPQY-----REVQCS-----EFN-	673	
Qy	131	DSNERTVFXHQSKAHHSKSWAIEGYOTROVSCTRSDGQNALSL-----CLQD	181	
Db	674	--NKDIGQGVASTNTHWVPKYANVAPNERCKLYC-RLSGSAFVLLLRKVDGTFCDEN	730	
Qy	182	SFPLTVQSCIMPXDC-----TSQSSNSPCKTCRSGSLLPF-----	219	
Db	731	GDDICVAGACMPAGCDHQLHSLRRDKCGCGDDSSCKVKGTFNEQGTFGYNEVMKIP	790	
Qy	220	FRSRNRNVHMAIGGCKECPLELEKACIVEGELLQCCPRYSWRTSEWKECOVSLLEQ-	278	
Db	791	AGSANDIRQKGNKNNKEDDNLVSLRA--ANGEFLN-----GHFQVSLARQOI	837	
Qy	279	--QD-----PHWHTGP-----VCGGGIQTREV---YCAQSVPAALRA	313	
Db	838	AFQDVTLEYSGSDAILERNGTPIRSYIVHVLSVGSHPDPISEYMTAAVENA-----	892	

314 KEYSRPVEKAL-----CVGPAP-----LPSQ- 334  
893 --VIRISSALIYLRVTDTECDRACRGOOSQKLMCLDMSTHRQSHDRNCQVLPKQOA 950  
335 --LCNIPCTDCIVSSWSAWGLCHENCHPEQKGFRTQRHVLHMGSTGPAHGHCHLVE 392  
951 TRMNCIDCSTRWITEDYS-----SCSAKSGGQKQORVSCVMEGRDQTPASEHLCD 1002  
393 --SVPCEDPMY-----RMLASEGI-CRPDGHKGLGHRILKAVCONDRGSDVSSLC- 442  
1003 RNSKPSDIASCYDCSGRKKNYGEWSTCETCSNGKMR--KSYCVDDSNRVEDSLC 1060  
443 -PVPPPERKSCIEPCRMDCVLSEWT--EWSQSQCSCNNKNSDGKQTRSRITLALAGE-- 497  
1061 REQKEATEECNRIPC-----PRWYGVHSECSRSKCDG---GVKWRHAQCLDAADRET 1110  
498 GGRPCPESQALQSHRLCNHDSQMLHETSMPWPCSEDTLVTALNATIGWGEATCGVI 557  
1111 HTRCGPAOT-QEH--CNEHAC--TWQFGVWSDCS-----AKGDDGV 1148  
558 QTRRVFCVSHGVQWTKRCPDSTRPETVRPCFLPKCKDCIVTAFSEWTFPCRMCO--- 613  
1149 QYRDANCTDRHRSVLPHRCLKMKIIT-KPCH---RESCPKYKLGWESQCSVSCEDGWS 1204  
614 -----AGNAT-----VKQSYRIIIEAANGQECPC----- 639  
1205 SRRVSCVSGNGTEVMSLCTASDRPASHOTCNLG-TCPEWRNTDWSACSVSOGIGHRR 1263  
640 --DTLYBERCEDVSLC-----PVYRWKPKQKSPCILVPESVMQGITSS 682  
1264 TTECIYREQSV-DASFCGDTKMPETSQCHLLPOTSKPWSHSPC-----S 1308  
683 EACGKGLQTRAVSC-----ISDNRSAEMME-CLKQTNMGPLLQVCECTVPCRE 729  
1309 VTCGSGIOTRSVSCTRGSEGTIVDEYFCDRTNRPRLKTCEDKTCDPGRVLQK---LOA 1364  
730 DCUFTAMS--KFTPCSTNCEATKRRRLQTKRKEKCDQ-----SDLYPLVET 777  
1365 DVFPFWATGFWTACSATCG-----NGTORULLKCRDHYRDLDPDECNHLDEVST 1415  
778 ELPCDFEISQFVGNWSDCILPBGREPHRGLRVQADSKCEGGLRPAVACSKDKNGRPV 937  
1416 RNCRLRDCSYWKAWECEPATCG--THVQOSRNVTCVSAEDGGRTILKQVDC-DVQKRET 1473  
838 DPSPCSSSGVIEQKCVIPCP-FDCKLSDM--SSWGSQSSCGIGVIRSKWLKEKPYNG 994  
1474 SARNCR-----LEPCPGEHIGSWIGDWSKCSASCSSGGRWRRSVSTSSCDET 1524  
895 RPKPLDLKNQVHEAVPCYSECNQ-----YSWVVEHWSCKINNELRSLRCGGGTQS 946  
1525 RK-PKM-----FDKNEELCPPLTNNSWQISPWTHC-----SVSCGGGVQR 1564  
947 RKIRCVNTADGEGAVDSNLNDEI-----PPEQSC-----SL 981  
1565 RKIWCEBVLGR-----KQDDIBCSIEIKPREQDCMPPCRSHYHNKTSASMTSL 1615  
982 MCPNECVMS-----EW--GLMSKCPQSCDPHTMORTRHLLRPSN-----S 1021  
1616 SSSNSNTTSSASASSLPIPVVSWQTSAMSACSACKGRGT-KERVVECVNPSLNVTVAS 1674  
1022 RTCAEDSQVQPC-----LLNENCFOFYNLTWSTCQSENAPCGQVYRTELLSCVSDGK 1077  
1675 TEC--DOTKXPEVEVRCTKHC--FRWKTITWSSCSVT-----CGRIRREVQCVRGN 1726  
1078 PVSMDQCEQHNLEKQPMISPLVECV-VNQCQLSGW--TAWTECSQTCGGRGMSRTRFI 1134  
1727 LVSPSECN-----PKTKLSNVANCFPVACPAYRNVNTPWSKCEBARGQKTRVHC 1779  
1135 IMPQGGRCPCPTBLTQET-----CPVTPC--YSWVLGNWSACKLEGDCGEGVQIRSLSC 1189  
1780 ISTSGKAAAPRMCELABAPTSIRSCTDSNCPYEWVPGDWQTC---SKSCGEGVQTRVRC 1836  
1190 --MYHSGSISHAAGREVDALCGEMPFQDSILKQLCSV-PCPGD---CHLT-----EWS- 1236

1837 RKXNFNSTIPIIWMLED-----EPAVPKEKCELFKPNESQTCELNCPDSEFKWSF 1888  
1237 -EWSTCELTG-----IDGRSPETV 1254  
1889 GPWGECKNCGQGIIRRRRVKVCVANDGRVERV 1920  
RESULT 2  
US-09-369-364A-13  
; Sequence 13, Application US/09369364A  
; Patent No. 6391810  
; GENERAL INFORMATION:  
; APPLICANT: Apte, Suneel  
; APPLICANT: Hurskainen, Tiina L.  
; APPLICANT: Hirohata, Satoshi  
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases  
; FILE REFERENCE: 26473/4007/10-30-00  
; CURRENT APPLICATION NUMBER: US/09/369,364A  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1882  
; TYPE: PRT  
; ORGANISM: Homo sapiens ADAMTS-9  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (468)  
; OTHER INFORMATION: Xaa = C  
; NAME/KEY: MOD RES  
; LOCATION: (521)  
; OTHER INFORMATION: Xaa = Y  
US-09-369-364A-13  
Query Match 8.0%; Score 657; DB 4; Length 1882;  
Best Local Similarity 21.2%; Pred. No. 3.5e-44;  
Matches 288; Conservative 123; Mismatches 421; Indels 526; Gaps 70;  
QY 201 WSS---NSPCKTCSGSLPLGPRSRNRVKNHMAIGGKCEPPELLEKEACIVEGELL--- 254  
DB 828 WNSHGPWACSKPCQ-----GERK-----KLVTRSDQLTVS 861  
QY 255 -QQCPRYSWRTSEKCEQVSLLEQQDPHWHVT-----GPVCGGGIQTREYVCAQSVPA 308  
DB 862 DQRCDELPPQGHITPCGTGC---DLRWHVASRSECSAQCGLGRTLDIYCAK----- 911  
QY 309 AALRAKEVSRPVEKALCVGPAPLPSQLCNTPCSTDCIVSSW--SAWGLCIHENCHEPQK 366  
DB 912 -YSRLDGKTEKVDGDFC--SSHPKPSN--REKSGECNTGGRYSATWEC----- 956  
QY 367 KGFRTRQRHVLMESTGPAGHCPHLVESPCEDPMPCYRWLASGICFPDGHKCGLGHRLK 426  
DB 957 -----SKSCDGGTQRR 968  
QY 427 AVQNDREDVSGSLCPVPPPPERKSC-EIPCRMDCVLSEWT--EWSQSQCSCNNKNSDG 483  
DB 969 AICVNTNRDVLDDSKCTHQEKVTIQRCEFP-----PQKSGDWSECLVTC-----G 1016  
QY 484 KQTRSRITLALAGE---GGKPCPPSQALQEHRLCNHDSQMLHETSMPWPCSEDTLVT 540  
DB 1017 KGHKHSQVQCFGEDRLNDRMCDPETKTSNQTQOPECAS--WQAGPWQCS----- 1067  
QY 541 LNAITGWNGEATCGVIGTQTRVFC-VKSHVQVQWTKRCPDSTRPETVRPCPLPKCKDC-- 597  
DB 1068 -----VTCGGYQLRAVKCIIGTYSVVDNDCNATRTPTDQDCLEP---SCHP 1114  
QY 598 -----IVTAFSEWTPCPRMCAQNAVTQKSRYRIIIEAANGQECPCDT 641  
DB 1115 PPAPESTRSTYSAPRTQWRFGSWTPCSATCGK-----TRMYVSCRDENG-----S 1162  
QY 642 LYEEREC-----EDVSLCPVYRWKPKQWSFCILVPESVMQGITGSSEACGKGLQTR 692

1163 VADESACATLPRVAKERCSVTPOGMKALDSSC-----SVTCQGRATR 1208  
693 AVSCLSDNRAEMMECLKQINGPLLVQECTVPCREDCTFTAWSKTPCTSTNCEATKSR 752  
1209 QVMCVNSDHWDRSEC--DDYIPETDQCSM-----SPC-----1242  
753 RRLQTKSRKEKCDSDL--YPLVETELCPDEFISQY-----GNWSDCILPEG 801  
1243 -----PQTPDGLAHPQNEIDYRPSASPSRTHVLGNQWRTGPGAC-----1287  
802 RREPHRGLRVAADSKCEGEGELFRVAVACDKNX-----RPVDPSCSSSGYIOEX 851  
1288 -----SSTCAGSQRRVVVQDENGXTANDCVERIKPDEGRACESG-----1328  
852 CVIPCPDCKLSDW--SWGSCSSCGIGVIRSKWLKEKPYNGRCPKLDLKNVHEA 909  
1329 ---PCP-----QWYAGNWGECTKCGGIRTR---LVVSORSNGERFP--DLSCBILDK 1374  
910 VPCYSECNOY-----SNVVEHWSCKINNELRLRSGGQTSRKIRCVNTADGEGGAVD 963  
1375 PPDREQCNTHACPHDAWSTGFWSSC-----SVSCGRGHKQRNVYCMK--DGSHLE 1424  
964 SNLQNDQDEIPETQSC--SLMCPNECVMSEM--GLWSKCPQSCDHTWQRTPHLLRLSLN 1020  
1425 SYCCKHLAKPHGHRKCRGRCP-----KWAGAWSCSVSMGRGVQORH---VGQOIG 1474  
1021 SRTCAEDSQVOPCLLNEN-----CFQOYNLTWSTCOLSENAPCGQGVTRLLSCV 1072  
1475 THKIARETECPNTPRESECECQGRCPPLYTWABEWQEC---TKTCGEGSRYKVVCV 1530  
1073 -----CSDGK--PVSMDQCEQHNLEKQPMQSIPLCECVNVCOLSWTA--WTEC 1118  
1531 DNKNVHGARDVSKRPVDRSCSLQ-----PC--EYV-----WTTGEMSEC 1571  
1119 SOTCGHGRMSRTRIIMPTQEGRCPCTELTOEKT--CP-----VTPCY 1162  
1572 SVTCGKYQKQLVSCSEIYTGKEN---YEYSQYITINCPGTQPPSVHPCLRECPVSAT 1627  
1163 VVLGNWSACKLEGGCGGQGVQIRISUSCMVHSGSISHAA-----1200  
1628 MRVGNWGSV---SCGVGMQORSVQCLTNEDQPSHLCHTLKPBERTKCRNVYNCCLPQ 1684  
1201 -----GRVEDALCE--MPFODSILKQCSVPCPDCHLTWSEWSCTELTCIDGR 1249  
1685 NCKEVRLKAGED--GEYFLMIRGKLKIFCA-----GMHSDHPKEVYT--LVHGDSE 1734  
1250 SF-ETVGRQSRRTFIIOFENQDSC-----PQOVLET-----1281  
1735 NPSEVVGHLNPTPCPNYNGSRDDCQCKRDYTAAGFSFQKIRIDLTSMQIITDLQFA 1794  
1282 -----RP---CTGGKCYHT-----WKASLWNN---NERTVNCORS-----D 1312  
1795 RTSEGHVPFATAGDCYSAKCPQGRFSINLYGTGLSLTESARWISQGNVAVSDIKKSPD 1854  
1313 GYNVTGGCSFQAPPAIROCIPACRKPFSYCTQGGVCG 1350  
1855 GTRVVGKC-----GGYCG 1867

## :RESULT 3

IS-09-800-729-89

Sequence 89, Application US/09800729

Patent No. 6605592

GENERAL INFORMATION:

APPLICANT: Ni et al.

TITLE OF INVENTION: 32 Human secreted proteins

FILE REFERENCE: P2044P1

CURRENT APPLICATION NUMBER: US/09/800,729

PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: PCT/US00/26013

PRIOR FILING DATE: 2000-09-22

PRIOR APPLICATION NUMBER: 60/155,709

PRIOR FILING DATE: 1999-09-24

NUMBER OF SEQ ID NOS: 217  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 89  
LENGTH: 1745  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-800-729-89

Query Match 5.8%; Score 477.5; DB 4; Length 1745;

Best Local Similarity 19.2%; Pred. No. 1.2e-29; Indels 489; Gaps 51;

Matches 221; Conservative 124; Mismatches 318;

201 WSWSPSCSTCRSG---SLLPGFRSRS---RNVKMAIGGKECEPEL---LKEAKIVG 251

Db 39 WGWSSCSRTCGGASYSLLRCLSSKSCBGRNIRYRTC--SNVDCPEAGDFRAQCSAHN 97

Qy 252 ELLOQCPRYSWITSEWKECOVSLLEQDPHHVHTGVPVCGGIGIOTREVYCAQSVAAAL 311

Db 98 DVKHGQFYEW-----LPVSNDP-----DNPCSLKC 123

Qy 312 RAKEVSRPVEKALCVGPAPLPQLCNIPGSTDCIVSSWSAMGLCIHENCHEPQKKGFRT 371

Db 124 QAKGTTLVVELA-----PKVLDGTRCYTSL-----DMCISGLC 157

Qy 372 QRHVLMESTGAGHCPHLVSVPCDDPMCYRWLASBGICFPDHGKCGLGHRLKAVCON 431

Db 158 -----QIVG---CDHQLGSTVKED-----NCGVNGDGGSTCLRVGOYQSLSA 198

Qy 432 DRGEDV-----SGSLCPVPPPPERKSCETPCRMDCVLSEWTEWSSCSQSCSNKNSDGK 484

Db 199 TKSDDDTVVAIPYCSRHIRLVKGPDLHYLETK--TLOGTKGENSLSTGTFLVDNSSVDFQ 257

Qy 485 QTSRITLALAGEGKPCPPSOALQOHLRLCNHSCNQL-----HWTSWGPCSEBET 536

Db 258 KFPDKELRMAG---PLTADFIVKIRNSGSADSTVQFIYQPIIHRWRETDFPPCS---310

Qy 537 LVLTALNATIGNGEANTCGVGIOTRRVFCVKSHVGOVMTKR-C---PDSRREPVRCELP 592

Db 311 -----ATCGGGYQJTSABCYDLRSNRVVADQYCHYYPENIKP---KPLQOE 353

Qy 593 CKKDCIVTAFSEWTPCPRMCOAGNATVKOSRYRIIIOEAAANGQECPTLYBERECEDVS 652

Db 354 CNLD-----PCPA-----361

Qy 653 LCPVYWKQKWSFCLILVPESVWQGTGSSEACGKLOTRAVSCISDDNR---SAEMME 708

Db 362 ---RWEATPWTAC-----SSCGGGIQRVAVSCVEEDIQGHVTSVEEWK 402

Qy 709 CLKQTNMPLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKSRRLQATCKSRKKEKQD 768

Db 403 CM-YTPKPIA-----QPCN-----416

Qy 769 SDLYPLVETELCPDEFISQPYGNWSDCILPEGRPHRGLRVQADSKCEGLRPRAVA 828

Db 417 -----IFDCPKLAQ---ENSPCTV-----TCGQGLRYRVVL 445

Qy 829 CSDKNGRPVDPSPCSSSGYIOEKVIP-----CPFDCKLSDW-----865

Db 446 CIDHRGHTGGCSPKTKPHKEECIVFTCYKPEKLPVEAKL--PWFQAOALEEGAAYS 504

Qy 866 -----SSWGSOSSCGIGVIR-----SKWLKEKPNY---GGRPCPLDLKNQ 905

Db 505 EEPFPIPKAWSACTVTCGVGTQVIRVCQVLLSFQSADLPIDECEGPKPA-----SOR 559

Qy 906 VHEAVPCYSECNQ-----YSWVVEHWSCKINNELRLRSCGGGTS 946

Db 560 ACYAGCSGEIPFNDEDTGLFGGLQDFDELWDYEGFTKC-----SSCGGQVQE 612

Qy 947 RKIRCYNVADGEGGAVDSNLNQDEIPE--TOSCSLMCFNECVMSWGLSKCPQSCDPH 1005

Db 613 AVVSLNKKQTR--PAENLCVTSRPPQLKSCNL-----DPCPA-----651

Qy 1006 TMQRTRHLLRPSLSNRTCAEDSQVQPCLLNENCFOYNLTWSTCOLSENAPCGQVYR 1065



```
Db 652 -----RWEICKWSPCSLT-----CGVGLQ 670
Y 1066 TRLLSC-----VCSGKPVSMDCQCHNLEKQPMSPICLVECVWNCQ 1108
Db 671 TRDVFCHLLSRMNETVILADELCRQKPESTVOACNR-----FNCP 712
Y 1109 LSGWTA-WTECSOTCGH-----GRMSRTRFIIMPTQ--GEGRCPOPTELTOEKTCP 1156
Db 713 PAWYPAQWQPCSRTCGGVQKREVLCKQRMADGSFLELPETCSASKPACQACKKDDCP 772
Y 1157 VPCYVWLNWNSACKLEGGDCGSGVQIRSLSC--NVHSGSISHAAGRVEDALGEMPFQ 1214
Db 773 -----SEWLLSDMTECST--SCEGTQTRSACRKLMTG-----LSTVNVNSTLCPLLPFS 821
Y 1215 DSILKQLCSVPC 1226
Db 822 SSIR-----PC 827

RESULT 4
JS-09-132-769-3
; Sequence 3, Application US/09132769A
; Patent No. 6525023
; GENERAL INFORMATION:
; APPLICANT: Motoo Yamasaki
; APPLICANT: Kenji Shibata
; APPLICANT: No. 6525023uo Hanai
; APPLICANT: Akiko Furiya
; APPLICANT: Kaoru Miyamoto
; TITLE OF INVENTION: NOVEL VASCULAR SMOOTH MUSCLE CELL GROWTH FACTOR
; FILE REFERENCE: 11078
; CURRENT APPLICATION NUMBER: US/09/132,769A
; CURRENT FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: HE19-218491
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 807
; TYPE: PRT
; ORGANISM: BOVINE
JS-09-132-769-3

Query Match 5.3%; Score 440.5; DB 4; Length 807;
Best Local Similarity 21.2%; Pred No. 4.2e-27;
Matches 219; Conservative 123; Mismatches 340; Indels 351; Gaps 50;

Y 376 VLMESTGPA-----GHCPLHVESVPCDDPMCYRWLASGII-----410
Db 7 LLRLSRGFALLALPLAVALAFSDETLDKVPKSEGVCSRIILRVQTRREGVTEFSLRVE 66
Y 411 CFPDHGKGLGHRI-----LKAVQNDRGEDVSGSLCFVPPPPERKSC 453
Db 67 GDPDFYKPGTSYRVTLSAAPPSPVFRGFTLALKENREGKEEDHAGT-----113
Y 454 EIPCRMDCVLSEWTEWSSCSQSNKNSDKQTRSRITILALAGEGKPCPPSOALQEHRL 513
Db 114 -----FOIIEEETQFWSNCPVAVTESTPRRTR-----142
Y 514 CNDSHCMLWETSPWGPCSEDLIVTALNATIGNGEATCGVIGIOTRRVFCVKSHVQVM 573
Db 143 -----IQVFWIAPPAG-----TGCVILKASI-----VQRRIYF--QDEGSLT 178
Y 574 TKRC-PDSTRPETVRPFLPKCKDCIVTAFSEWTPCPRMCOAGNATVKSRYRIIIEAA 632
Db 179 KKLCEQDSTFDGVTDKPILOC-----CACGTA-----KYRLTFY--G 213
Y 633 NGGCECDPTLYEECECDVSLCPYRWKPKQNSPCLVPESVWQGTGSEACG-----686
Db 214 NWSEKTHPKDYPR-----ANHWSAIIIGSHSKNYVLYWEY 248
Y 687 KGLQTRAVSCISDDNRSANMECLKQTNMPLLVQECTVPCREDCTFTAWSKF--TPCSTN 745
```

```
Db 249 GGVASEGVKQVAELSGSVKWEIRQQSDVLTV-----IKAKQWPAWQPLNVRAAPS 302
Y 746 CEATKSRRL-----TKSRKKEKCCODSDLYPLVETELCPCEDFISQPYG 791
Db 303 AEFSDVTRHLMBSFLTMGSPSPDNVNLGSAEDLCTKECGWVKVQVQDLIPWDA-----G 356
Y 792 NWSICILPEGRREPHRLRVQADSKCEGELRPRVACSDKNGRPVDPSPFCSSGGYIQ--849
Db 357 TDS-----GVTVESFNKPTIPOEKIRPLTSLD--HPQSPFYDDEGGSTQV 400
Y 850 -----EKV-VIP-----CPFDCLSDWSWGSCSSS--CG 876
Db 401 ARVIERIARKGQCINIVPDNDVIDADLAPBKDEDDTPTCTIYSNWSWSPSACSSSTCD 460
Y 877 IGVRIRSKWLKEKPYNGRPRCPKLDLKNQVHEAVPCV-----SECNQYSWVVEHWS 927
Db 461 XGEMRQRMKQAQ-LDLVFCP-----DTQDFQPCMGPCSDSDGSTCTMSEWIT--WS 511
Y 928 SKINNELSLRCGGTQSRKIRCVNTADGEGAVDSNLCNQDEIP-PETQSCSL--MC 983
Db 512 PC-----SISCGTGRSRE-RYVKQFPEDG-----SVCT--LPTETEKTVAEES 555
Y 984 PNECVMSWGLWCKPCQSCDPHTMORRTRHLLRPLSLNSRTCAEDSOVQPCLLNENCFQFQ 1043
Db 556 PSSCLTTEWGEWDECSATCGMGKKRHWKMSPADGSMCKAETSQAECMPEE-CHTIP 614
Y 1044 YNL---TENSTQLSENAPCGQVRT--RLLSVCVSDGKPVSMDCQEQHNLKPKQMSIP 1098
Db 615 CLLSLWSEWSDCSVT---CGKGNRTQRMLKSLAELG-----DCNE-ELEQVEKCMPL 663
Y 1099 CLVECVNCSLGSWTATECSOTCGHGRMSRTRFIIMPTQSGEGRPCPTLTOEKTCPVT 1158
Db 664 ---ECPIDCELTEWSQWSECKNSCGK-GEMIRTMQMEPFQGTGTCF-ETVORKKCRIR 718
Y 1159 PCSWVLGNWSACKLEGGDCGEGVQIRSLSCMVHSGSISHAAGRVEDALGEMPFQDSIL 1218
Db 719 KC---LRNPSIQNLWRARE-----SRRLAESDGDQ-----751
Y 1219 KQLCSVPCGDCHLTEWSEWSTCELTICIDG--RSFETVGRQSRRTFILOSFENQDSCPQ 1276
Db 752 -----FPG-CRMRPTAWSECTKLGCGGIQIRYMTVKRFSKFSOF-----TSCKD 795
Y 1277 QVLETRPCCTGGKC 1289
Db 796 K-KEIRACNVHPC 807

RESULT 5
US-09-132-769-1
; Sequence 1, Application US/09132769A
; Patent No. 6525023
; GENERAL INFORMATION:
; APPLICANT: Motoo Yamasaki
; APPLICANT: Kenji Shibata
; APPLICANT: No. 6525023uo Hanai
; APPLICANT: Akiko Furiya
; APPLICANT: Kaoru Miyamoto
; TITLE OF INVENTION: NOVEL VASCULAR SMOOTH MUSCLE CELL GROWTH FACTOR
; FILE REFERENCE: 11078
; CURRENT APPLICATION NUMBER: US/09/132,769A
; CURRENT FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: HE19-218491
; EARLIER FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 807
; TYPE: PRT
; ORGANISM: human
US-09-132-769-1

Query Match 5.3%; Score 434.5; DB 4; Length 807;
```

Best Local Similarity 21.7%; Pred. No. 1.3e-26;  
Matches 216; Conservative 120; Mismatches 342; Indels 317; Gaps 50;

391 VESVPCEDPMCYRWLASGEI-----CFPDHGKGLGHRILKAVCONDRGEDV 437  
34 LDKVPKSEGYCSRIILRAQGTREGYTEFSLRVEGDPDFYKPGTSYRV----- 80  
438 SGLSCPVPPEPERKSCIEIPCMDCVLSEWSSCSQCSNKNKSGKQTSRTILALAGE 497  
81 --TLGAAPSYFPGFTLIALR-----ENREGDKEEDHAGTFOIIDEE 120  
498 GKG---PCPPSOALQEHRLCNDHSCMLHWETSPMGPCSEDTLVLTALNATIGWGEATCG 554  
121 ETQFMNSCPVAVTESTPR---RRTIQVFWIAPPAG-----TGCVILKASI----- 163  
555 VGIOTRPFVCKSHVQVMTKRC-PDSTRPETHVPCFLPKCKDCIVTAFSEWTPCPRMQ 613  
164 --VQKRIIYF--QDEGSLTKLCEQDSTFDGVTDKPILDC-----CA 201  
614 AGNATVKQSYRIIIEQAANGGQCEPDTLYEBRECEDEVSLCPVYRWKPKQWSPCILVPES 673  
202 CGTA-----KYRLTFY--GNWSEKTHPKDYPR-----AN 229  
674 VMGIGTGSSEACG-----KGIQTRAVSCISDDNRSAMWECCLKQTNMPLLVQECTVPC 727  
230 HWSAIIIGGSHSKNYVLWEYGGVASEGVKQVABLGSFVKMEEBIRQOSDEVLTV-----I 283  
728 REDCTFTAKSKF--TPCSTNCEATKSRRL-----TGKSRKKEKQDSDLY 772  
284 KAKAQPAPQPLNVRAPSAEFSVDRTHLMGFLTMGSPDPNVLGSAEDLCTKECGWV 343  
773 PLVETELCPDCEFIISQPYGNWSDCILPEGRREPHRGLRVQADSKCEGGLRFRVACSDK 832  
344 QKVVDLIPWDA-----GTDS-----GVTYESPNKPTIPEKIRPLTSLD- 383  
833 NGRPVPDPSCSSSGYIQ-----BKC-VIP-----CPF 858  
384 --HPOSPFYDPGEGSITQVARVIERIARKGQCNIVPDNDVDDIVADLAPEKDEDDTPE 441  
859 DCKLSWSSWGSCSSS-CGIGVRIKSKMLKEPYNGRCPKPLDLKNQVHEAVPCY----- 913  
442 TCIYSNWSWPSACSSSTCDKKGEMRQMLKAQ-LDLSVPCP-----DTQDFQPCWGPC 494  
914 -----SECNQYSWVHEHWSCKINNELSLRCGGTQGRKRCVNTADGEGAVDSNLN 968  
495 SDEGSGTCTMSEWIT--WSPC-----SISCGMGRSRE-RYVKQFPEDG-----SVCT 539  
969 QDEIPPE-----TQSCLMCPNECVMSWGLWSKQPCSDPHMTQRTTRHLLRPSLNS 1021  
540 ---LPTEEKTEKTVNECS---PSSCLMTWGEWDECSATCGMKKRRHMKMNPADGS 593  
1022 RTCAEDSQVQPCLLNENCFQOYNL---TEWSTCQLSENAPCGQVRT--RLLSCVCSDG 1076  
594 MCKAETSQAELKQWPE-CHTIPCLLSPNSEDSCSVT---CGKMRTRQMLKSLAELG 648  
1077 KPVSMDQCEHNLKXPORMSIFCLVECVNVCOLSWTANTECSQTCGGRMSRTRFIIM 1136  
649 -----DCNE-DLEQVEKQWMLP--ECPDCLTELTEWSQWSEKNSCGK-GHVIKTRMIQM 697  
1137 PTQGEGRPCPTLTQKTCPTVPCYSWVLGNWSACKLEGDCGEGVQIRSLSCMVHSGSI 1196  
698 EPQFGGAPCP-ETVQPKKCRKCK--LRNPSIQKPRWREARE----- 736  
1197 SHAAGREVALGEMFPQDSILKQLCVPQDCHLTWSEWSTCELTCIDG--RSFETV 1254  
737 SRRSEQLKEESEGEQ-----FPG-CRMRPWTAMSECTKLCGGGIQERYMTV 781  
1255 GROSRSRTIIQSFENQDSCPOOVLETRPCTGGK 1289  
782 KXRFKSSQF-----TSCKDK-KEIRACNVHPC 807

RESULT 6

US-09-640-173-186  
; Sequence 186, Application US/09640173  
; Patent No. 6613515  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Stolk, John A.  
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND  
; TITLE OF INVENTION: METHODS OF USE THEREFOR  
; FILE REFERENCE: 210121.484C2  
; CURRENT APPLICATION NUMBER: US/09/640,173  
; CURRENT FILING DATE: 2000-08-15  
; NUMBER OF SEQ ID NOS: 196  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 186  
; LENGTH: 807  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-640-173-186

Query Match 5.3%; Score 434.5; DB 4; Length 807;  
Best Local Similarity 21.7%; Pred. No. 1.3e-26;  
Matches 216; Conservative 120; Mismatches 342; Indels 317; Gaps 50;

391 VESVPCEDPMCYRWLASGEI-----CFPDHGKGLGHRILKAVCONDRGEDV 437  
34 LDKVPKSEGYCSRIILRAQGTREGYTEFSLRVEGDPDFYKPGTSYRV----- 80  
438 SGLSCPVPPEPERKSCIEIPCMDCVLSEWSSCSQCSNKNKSGKQTSRTILALAGE 497  
81 --TLGAAPSYFPGFTLIALR-----ENREGDKEEDHAGTFOIIDEE 120  
498 GKG---PCPPSOALQEHRLCNDHSCMLHWETSPMGPCSEDTLVLTALNATIGWGEATCG 554  
121 ETQFMNSCPVAVTESTPR---RRTIQVFWIAPPAG-----TGCVILKASI----- 163  
555 VGIOTRPFVCKSHVQVMTKRC-PDSTRPETHVPCFLPKCKDCIVTAFSEWTPCPRMQ 613  
164 --VQKRIIYF--QDEGSLTKLCEQDSTFDGVTDKPILDC-----CA 201  
614 AGNATVKQSYRIIIEQAANGGQCEPDTLYEBRECEDEVSLCPVYRWKPKQWSPCILVPES 673  
202 CGTA-----KYRLTFY--GNWSEKTHPKDYPR-----AN 229  
674 VMGIGTGSSEACG-----KGIQTRAVSCISDDNRSAMWECCLKQTNMPLLVQECTVPC 727  
230 HWSAIIIGGSHSKNYVLWEYGGVASEGVKQVABLGSFVKMEEBIRQOSDEVLTV-----I 283  
728 REDCTFTAKSKF--TPCSTNCEATKSRRL-----TGKSRKKEKQDSDLY 772  
284 KAKAQPAPQPLNVRAPSAEFSVDRTHLMGFLTMGSPDPNVLGSAEDLCTKECGWV 343  
773 PLVETELCPDCEFIISQPYGNWSDCILPEGRREPHRGLRVQADSKCEGGLRFRVACSDK 832  
344 QKVVDLIPWDA-----GTDS-----GVTYESPNKPTIPEKIRPLTSLD- 383  
833 NGRPVPDPSCSSSGYIQ-----BKC-VIP-----CPF 858  
384 --HPOSPFYDPGEGSITQVARVIERIARKGQCNIVPDNDVDDIVADLAPEKDEDDTPE 441  
859 DCKLSWSSWGSCSSS-CGIGVRIKSKMLKEPYNGRCPKPLDLKNQVHEAVPCY----- 913  
442 TCIYSNWSWPSACSSSTCDKKGEMRQMLKAQ-LDLSVPCP-----DTQDFQPCWGPC 494  
914 -----SECNQYSWVHEHWSCKINNELSLRCGGTQGRKRCVNTADGEGAVDSNLN 968  
495 SDEGSGTCTMSEWIT--WSPC-----SISCGMGRSRE-RYVKQFPEDG-----SVCT 539  
969 QDEIPPE-----TQSCLMCPNECVMSWGLWSKQPCSDPHMTQRTTRHLLRPSLNS 1021  
540 ---LPTEEKTEKTVNECS---PSSCLMTWGEWDECSATCGMKKRRHMKMNPADGS 593  
1022 RTCAEDSQVQPCLLNENCFQOYNL---TEWSTCQLSENAPCGQVRT--RLLSCVCSDG 1076

Db 594 MCKAETSQAEBKMMPE-CHTIPCLLSPWSEWSDCSVT-----CGKGMRTQRMLKSLAELG 648  
2Y 1077 KPVSMQCEQHNLEKQPMISIPCLVECVNQLSGWTAWTECSQTCGHGGRMSTRFIIM 1136  
Db 649 -----DCNE-DLEQVEKCMPL--ECPIDCELTEWSQSECNKSCGK-GHVIRTRMIQM 697  
2Y 1137 PTQGEGRPCPTLTQKTCPTVPCYSVNLGNWSACKLEGDCGEGVQIRSLSCMVHSGSI 1196  
Db 698 EPQFGGAPCP-ETVQRKKCRKRC-----LRNPSIQKPRWREARE----- 736  
2Y 1197 SHAAGRVEDALCCEMPFQDSILKQLCSVPCGDLHTEWSEWSTCELTICDG--RSFETV 1254  
Db 737 SRRSEQLKEESEGEQ-----FPG-CRMRPMTAWSECTKLCGGGIQERYMTV 781  
2Y 1255 GROSRSRTFIQSPENODSCPOQVLETRPCTGGKC 1289  
Db 782 KRFKSSQF-----TSCKDK-KEIRACNVHPC 807  
  
RESULT 7  
US-09-713-550-186  
; Sequence 186, Application US/09713550  
; Patent No. 8617109  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; THERAPY AND DIAGNOSIS OF OVARIAN CANCER  
; FILE REFERENCE: 210121.484C4  
; CURRENT APPLICATION NUMBER: US/09/713.550  
; CURRENT FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 205  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 186  
; LENGTH: 807  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-713-550-186  
  
Query Match 5.3%; Score 434.5; DB 4; Length 807;  
Best Local Similarity 21.7%; Pred. No. 1.3e-26;  
Matches 216; Conservative 120; Mismatches 342; Indels 317; Gaps 50;  
  
QY 391 VESVPCDDPMCYRLASEGI-----CFPDHKGCGLGHRLKAVQCONDRGEDV 437  
Db 34 LDKVPKSEGYCSRLRAQGTREGEYTBFSLRVEGDDPFYKPGTSYRV----- 80  
QY 438 SGSLCPVPPPPPERKSCIPCRMDCVLSEWTEWSSCSQSCSNKNSDGKQTRSRITILALAGE 497  
Db 81 --TLSEAAPPYFRGFTLIALR-----ENREGDKEEDHAGTFQIIDE 120  
QY 498 GKG---PCPPSQALQEHRLNDHSCMLHETSPWGPCSEDTLVTALNATIGNGEATCG 554  
Db 121 ETQFMSNCPVAVTESTPR--RRTRIQVFIAPAG-----TGCVLKASI----- 163  
QY 555 VGIOTRRVFCVSHVGQVMTKRC-PDSTRPETVRPCFLPKCKDCIVTAFSEWTPCPRMCQ 613  
Db 164 --VQKRIIYF--QDEGLTKLCEQDSTFGVTDKPILOC-----CA 201  
QY 614 AGNATVQSRVRIIIOBAANGQCEPPTLYEERECEDVSLCPVVRKPKQKMSPCILLVPS 673  
Db 202 CGTA-----KYRLTFY--GNWSEKTHPKDYPR-----AN 229  
QY 674 VWOGITGSSEACG-----KGLQTRAVSCISDDNRSAMMECLKQTMGPLLVQECTVPC 727  
Db 230 HWSAIIIGSHSKNVLYEYGYAGSEGVQVAELGSPVWMEERIRQSGDEVLTV-----I 283  
QY 728 REDCTFTAWSKF--TPCSTNCEATKSRRLQ-----TGSKRKKCEKQDSLDY 772  
Db 284 KAKAQWPAQWPLNVARAAPSAEFSVDRTHILMSFLTMWGPSPDNVGLSAEDLCTKCGWV 343  
QY 773 PLVETELCPCEDEFISQPGNWSDCILPEGRREPHRGLRVQADSKECGEGLRFFAVACSDK 832

Db 344 QKVVDLIPWDA-----GTDS-----GVTESPNKPTIQEKIRPLTSLD- 383  
QY 833 NGRPVDPFSCSSSYIQ-----EXC-VLP-----CPF 858  
Db 384 --HPOSFPYDEGSIQTQVARVIERIARKGEQCNIVPDNVDDIVADLAPEEKDDTPE 441  
QY 859 DCKLSDHSSWGSSESS--CGIGVIRSRKWLAKPVGGRPCPKDLKQVHEAVPCY---- 913  
Db 442 TCIYSNWSPWACSSSTCDKGRQRMLRAQ--LDLSVPCP-----DTQDFQCMGPGC 494  
QY 914 -----SECNQYSWVVEHWSCKINNELRLRCGGGTQSRKIRCVNTADGGGAVDSNLGN 968  
Db 495 SDEGSTCTGSEWIT--WSPC-----SISCGMGMSRE-RYVKQPPEDG-----SVCT 539  
QY 969 QDEIPPE-----TQCSLMLCPNECVMSWGLWSKCPQSCDPTMORRTEHLLRPSLNS 1021  
Db 540 ---LPTMEMEKTVNEBES---PSSCLMTWGEWDECSATCGMKMKRHRMKNPADGS 593  
QY 1022 RTCAEDSQVQCLINENCFOFYNL---TEMSTQLSENAPCGQVRT--RLLSVCVSDG 1076  
Db 594 MCKAETSQAEBKMMPE-CHTIPCLLSPWSEWSDCSVT-----CGKGMRTQRMLKSLAELG 648  
QY 1077 KPVSMQCEQHNLEKQPMISIPCLVECVNQLSGWTAWTECSQTCGHGGRMSTRFIIM 1136  
Db 649 -----DCNE-DLEQVEKCMPL--ECPIDCELTEWSQSECNKSCGK-GHVIRTRMIQM 697  
QY 1137 PTQGEGRPCPTLTQKTCPTVPCYSVNLGNWSACKLEGDCGEGVQIRSLSCMVHSGSI 1196  
Db 698 EPQFGGAPCP-ETVQRKKCRKRC-----LRNPSIQKPRWREARE----- 736  
QY 1197 SHAAGRVEDALCCEMPFQDSILKQLCSVPCGDLHTEWSEWSTCELTICDG--RSFETV 1254  
Db 737 SRRSEQLKEESEGEQ-----FPG-CRMRPMTAWSECTKLCGGGIQERYMTV 781  
QY 1255 GROSRSRTFIQSPENODSCPOQVLETRPCTGGKC 1289  
Db 782 KRFKSSQF-----TSCKDK-KEIRACNVHPC 807  
  
RESULT 8  
US-09-800-729-217  
; Sequence 217, Application US/09800729  
; Patent No. 6605592  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: P2044P1  
; CURRENT APPLICATION NUMBER: US/09/800.729  
; CURRENT FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: PCT/US00/26013  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: 60/155,709  
; PRIOR FILING DATE: 1999-09-24  
; NUMBER OF SEQ ID NOS: 217  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 217  
; LENGTH: 1059  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-800-729-217  
  
Query Match 5.2%; Score 429; DB 4; Length 1059;  
Best Local Similarity 22.3%; Pred. No. 5.2e-26;  
Matches 263; Conservative 112; Mismatches 382; Indels 424; Gaps 73;  
  
QY 418 CGLG-----HRLKAVQONDRGEDVSGSLCPVPPPERKSCIEPCRD-----CVLSWTE 468  
Db 39 CGGVSVRQLRCLTSKCS---GESVRFVK-----AQKTESKSLRARDTICGEELVS 89  
QY 469 WSSCSQSCSNKNSDGKQTRSRITILALAG-----EGKPCPPSQALQEHRLCNDHSC-- 519  
Db 90 RQCEVVC-----RSR-----LTGANFLWRVDDGTPC---QAATSRVAVCSKGCQI 132





CT-US93-03164-10

Sequence 10, Application PC/TUS9303164

GENERAL INFORMATION:

APPLICANT: Jessell, Thomas M

APPLICANT: Klar, Avihu

TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A

TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESS: Cooper &amp; Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/03164

FILING DATE: 19930402

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: White, John P

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 40028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 977-9550

TELEFAX: (212) 664-0525

TELEX: 422523 COOP UI

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 807 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

CT-US93-03164-10

Query Match

Best Local Similarity 5.2%; Score 428.5; DB 5; Length 807;

Matches 133; Conservative 64; Mismatches 158; Indels 107; Gaps 27;

17 849 QEKVIPCDFCDKLDWSWGCSSS-CGIGVIRSKILKEKPYNGGRPCPKLDLKNQVH 907  
18 432 EKDEDDTPTETCIYGNWSPWASCSSTCEKGRMRQRLKAO-LDLSVPFCP-----DTQ 484  
19 908 EAVPCV-----SECNQYGVVVEHWSCKINNELSLRCGGGTQGRKIRCVNTADGE 958  
20 485 DFQPCMGPCSDDEGDTCTWSEWIT--NSPC-----SVSCGMGRSRE-RYVKQFFED 534  
21 959 GGAVDNLCNQDEIP-PETQSSCL---MCPNCEVMSWGLKSCQSCDPHTMQRTTHL 1014  
22 535 G-----SVC---MLPTEETKCTVNEECSPSSCLVTWGEWDDCSATCGMGKKGRHVMK 586  
23 1015 LRPLNSRITCAEDSOVQPCLLNENCFQFQYNI---TEWSTCOLSENAPCGQGVRT--RL 1069  
24 587 MFAFGSNCKATSKAQEKMWPE-CHTIPCLLSPWSEWSDCVT-----CGKMRTRQRL 641  
25 1070 SCVSDGKPVMSQDCEQHNLEKPRQMSIPCLVECVNVCOLSGWTATWTCSTQCGHGRMS 1129  
26 642 KSLAELG-----DCNE-DLEQAEKMLP---ECPIDCELSWSQWSECNKSCGR-GHMI 690  
27 1130 RTFIIIMPQGRPCPTLTQKTCPTVPCYSWLGNWSACKLEGGDCGEGVQIRSLSC 1189  
28 691 RRTQMEPQFGAPCP-ETVQKKCRARKC-----LRSPSIQKLWRARE----- 736  
29 1190 MVHSGSISHAAGRVEDALCGEPFQDSILKQLCSVPCGDCHLTWSEWSTCELTCDIG- 1248

Db 737 -----SRRSEQLREESDGEQ-----FPG-CRMRPWTAWSECTKLCGGGI 774

Qy 1249 -RSFETVGRQSRRTIIQSFENQDSCPOQVLETRPCTGKGC 1289

Db 775 QERYMTVKKRFKSSQF-----TSCKDK-KEIRACNVHPC 807

RESULT 13

US-07-862-021B-12

Sequence 12, Application US/07862021B

Patent No. 5279966

GENERAL INFORMATION:

APPLICANT: Jessell, Thomas M

APPLICANT: Klar, Avihu

TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A

TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper &amp; Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/862,021B

FILING DATE: 19920405

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: White, John P

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 40028

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 977-9550

TELEFAX: (212) 664-0525

TELEX: 422523 COOP UI

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 802 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-862-021B-12

Query Match 5.1%; Score 419; DB 1; Length 802;

Best Local Similarity 23.8%; Pred. No. 2.3e-25;

Matches 186; Conservative 100; Mismatches 277; Indels 220; Gaps 44;

Qy 605 WT--PCPRMCOAGNATVKOSRY-----RIIQBAANGQCEPDTLYEERECEV 651  
Db 142 WTAPPTGTGVILKASIVQRIYIFQDEGSLTKRICEQDSASEGVTDKPTL-----DC 194  
Qy 652 SLCPV--YR-----WKPKWSPCIIIVPESVWQGITGSSEACG-----KGLQTRAVSCI 697  
Db 195 CACGTAKYRLTFYGNWSEKTHPKDFPRRTNHWSAIIIGSSHKNYILWEYGVYASEGVKQV 254  
Qy 698 SDDNRSNEMECILKQINGNPLIVQECTVPCREDCTFTANSKF-TPCSTNCEATKSRRL 756  
Db 255 AELGSPVMEERIEQSDSEVLTV-----IKAKAQPAMQPLNVRAPSAEFSVDRHRL 308  
Qy 757 -----TGKSRKKEKQSDILYVETELPCDDEFISQPYGNWSDCILPEGR 802  
Db 309 MSFLTMLGPSDNVGLSADLCTKCGWVKVQDLIPWA-----GTDS----- 354  
Qy 803 REPHRGLRVQADSKCEGEGRLFRFVACSKDNKGRPVPDPSPCSSSGYIQ----- 849  
Db 355 -----GVTYESPNKPTVPOEKIRPLTSLD---HPQSPFYDPEGSGIKLVAVVLERIARK 406

850 -EKC-VIP-----CPFDCKLSWMSWSSSS--CGIGVIRSRKMLK 887  
407 GEQCNFVNDIDIVADLAPEKEEDTPTETCIYSNWSWSSACSSSTCEKGRMRQRLK 466  
888 EKPNGGRPCPKLDLKNQVHEAVPCY-----SECNQYSWVVEHSSCKINNELSL 938  
467 AQ-LDLSVPCP-----DTQDFQPCMGPGCSDEGTCMMSDWT--WSPC-----SV 510  
939 RCGGTSQSRKIRCVNTADGEGGAVDSNLQNDRIIP-PETQSCSL---MCPNECVMSWGL 994  
511 SCGMGTRSRE-RYVKQPPEDG-----SMC---KVPTETEKCIWNEBESSCLVTEWGE 561  
995 WSKCPQSCDPTHMQRTRHL-LRPSLNSRTCAEDSQVQCLLNENCFQYNL---TEWS 1050  
562 WDECSASCGT-GMKRRHRMKTADGSMCKAETTEAKCMMP-CHTIPCLLSFNSW 619  
1051 TCQLSENAPCGQGVRT--RLLSVCVSDGKPVMSDQCHNLEKPFQMSIPCLVECVNQC 1108  
620 DCSVT-----CGKGRTRQRLKSAELG-----DCNE-ELEQAEKMLP---ECPIDCE 665  
1109 LSGWTAWTECSQTCGHGRMSRTRFIIMPTQGEGRPCPTLTQEKTCPTVPCYSWVLGNW 1168  
666 LTEMSONWSECNSTCGK-GHMIRTRMKIEPQFGGTACP-ETVQRTKCRVRKCLR----- 717  
1169 SACKLEGDCGEGVQIRLSLSCMVHSGSISHAAGRVEDALCGEMPFDOSILKOLCSVPCPG 1228  
718 -----GPGMEKR-----RWKEA--REKRSQAQKNIDNEQYP- 748  
1229 DCHLTWSENSTCLTCLDIDG--RSFETVGRQSRRTFIQSFNQDSQCPQVLETRCTG 1286  
749 VCRLLKPWTAWTECSTLGGGQIOBYMYMKRSKSTQF-----TSCKDK-KELRACNV 799  
1287 GKC 1289  
800 HPC 802

## RESULT 14

US-08-313-288B-12  
; Sequence 12, Application US/08313288B  
; Patent No. 5750502  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M. and Avihu Klar  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A  
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/313,288B  
; FILING DATE: January 5, 1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0526  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 802 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-313-288B-12  
Query Match 5.18; Score 419; DB 1; Length 802;  
Best Local Similarity 23.8%; Pred. No. 2.3e-25; Indels 220; Gaps 44;  
Matches 186; Conservative 100; Mismatches 277;  
605 WT--PCPRMCOAGNATVKQRY-----RIIQEAAAGQCEPDTLYBERECEDV 651  
142 WTAPPTGTCGVILKASIVQKRIIVQDEGSLTKRICBQDSASEGVTDKPTL-----DC 194  
652 SLCPV--YR-----WPKQWSPCLILVPESVWQIGTGSSEACG-----KGLQTRAVSCI 697  
195 CACGTAKTRUTFYGNWSEKTHPKDFPRTHNSAILGSSHKNYILWEYGGYASEGVKQV 254  
698 SDDNRSAAEMCEKQTNGMPLLVQECTVPCREDCTFTAMSKF-TPCSTNCEATKSRRL 756  
255 AELGSPVKWEEIRFQSDDEVITV-----IKAKAQWPAWQPLNVAAPSAEFSVDRHRL 308  
757 -----TGSRKKEKQDSOLYPLVETELCPDCEFIQYGNWSDCILPEGR 802  
309 MSFTMLGPSDMNVGLSAEDLCTKDCGWYQVQDLPWDA-----GTDS----- 354  
803 REPHRLRVQADSKECGEGLRFRVACSDKNGRPVDPSPFCSSSGYIQ----- 849  
355 -----GVYESPNKPTVPOEKIRPLTSLD--HPQSPFYDPEGSSIKLVARVLERIARK 406  
850 -EKC-VIP-----CPFDCKLSWMSWSSSS--CGIGVIRSRKMLK 887  
407 GEQCNFVNDIDIVADLAPEKEEDTPTETCIYSNWSWSSACSSSTCEKGRMRQRLK 466  
888 EKPNGGRPCPKLDLKNQVHEAVPCY-----SECNQYSWVVEHSSCKINNELSL 938  
467 AQ-LDLSVPCP-----DTQDFQPCMGPGCSDEGTCMMSDWT--WSPC-----SV 510  
939 RCGGTSQSRKIRCVNTADGEGGAVDSNLQNDRIIP-PETQSCSL---MCPNECVMSWGL 994  
511 SCGMGTRSRE-RYVKQPPEDG-----SMC---KVPTETEKCIWNEBESSCLVTEWGE 561  
995 WSKCPQSCDPTHMQRTRHL-LRPSLNSRTCAEDSQVQCLLNENCFQYNL---TEWS 1050  
562 WDECSASCGT-GMKRRHRMKTADGSMCKAETTEAKCMMP-CHTIPCLLSFNSW 619  
1051 TCQLSENAPCGQGVRT--RLLSVCVSDGKPVMSDQCHNLEKPFQMSIPCLVECVNQC 1108  
620 DCSVT-----CGKGRTRQRLKSAELG-----DCNE-ELEQAEKMLP---ECPIDCE 665  
1109 LSGWTAWTECSQTCGHGRMSRTRFIIMPTQGEGRPCPTLTQEKTCPTVPCYSWVLGNW 1168  
666 LTEMSONWSECNSTCGK-GHMIRTRMKIEPQFGGTACP-ETVQRTKCRVRKCLR----- 717  
1169 SACKLEGDCGEGVQIRLSLSCMVHSGSISHAAGRVEDALCGEMPFDOSILKOLCSVPCPG 1228  
718 -----GPGMEKR-----RWKEA--REKRSQAQKNIDNEQYP- 748  
1229 DCHLTWSENSTCLTCLDIDG--RSFETVGRQSRRTFIQSFNQDSQCPQVLETRCTG 1286  
749 VCRLLKPWTAWTECSTLGGGQIOBYMYMKRSKSTQF-----TSCKDK-KELRACNV 799  
1287 GKC 1289  
800 HPC 802  
RESULT 15  
PCT-US93-03164-12  
; Sequence 12, Application PC/TUS9303164  
; GENERAL INFORMATION:  
; APPLICANT: Jessell, Thomas M  
; APPLICANT: Klar, Avihu  
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A

TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Cooper & Dunham  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03164  
FILING DATE: 19930402  
PRIOR APPLICATION DATA:  
CLASSIFICATION:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 40028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 664-0525  
TELEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 802 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
CT-US93-03164-12

Query Match 5.1%; Score 419; DB 5; Length 802;  
Best Local Similarity 23.8%; Pred. No. 2.3e-25;  
Matches 186; Conservative 100; Mismatches 277; Indels 220; Gaps 44;

17	605	WT--PCPRMCQAGNATVKQSRV-----RIIQEAANGQCEPDITLYEECEEDV	651
18	142	WTAPPTGTGCVILKASIVQKRIIYFQDEGLTKRICEQDSASEGVTDKPTL-----DC	194
19	652	SLCPV--YR-----WKPKWSPCLVPESVWQGITGSSEACG-----KGLQTRAVSCI	697
20	195	CACGTAKYRLTFYGNWSEKTHPKDPRRTNHSALIGSSHSHKYNILWEYGVYASEGVKQV	254
21	698	SDNRSANWECLKQTNMPLLVQECTVPCREDCTFTAWSKF-TPCSTNCEATKSRRLQL	756
22	255	AEIGSPVMEIEIRQOSDEVLTV-----IKAKAQPAPQPLNVRAAPSAEFSVDRHRL	308
23	757	-----TGKSRKKEKQDSLDLYPLVETELCPCEBFIQPYGNWSDCILPEGR	802
24	309	MSFLTMLGSPDNWVGLSAEDLCTKCGWQVQVQDLIPWA-----GTDS-----	354
25	803	REPHRLRVQADSEKCEGELRFAVACSKNKRPRVDPSPFCSSSGYIQ-----	849
26	355	-----GVTYESPNKPTVQEKIRPLTSLD---HPQSFFYDPEGGSIKLVARVVLRIARK	406
27	850	-EKC-VIP-----CPPDCKLSDWSSWGSCSSS-CGIGVIRSRKMLK	887
28	407	GEQCNFVDNIDIVADLAPEEKEEDDTTETCIYNSWSPWASACSSSTCEKGRMRQMLK	466
29	888	EKPYNGRCPCKDLKLNQVHEAVPCY-----SECNQYSWVVEHSSCKINNELRSL	938
30	467	AQ-LDLSPVCP-----DTQDFQCMGPGCSDEBGTGCMWSDWIT--NSPC-----SV	510
31	939	RCGGGTQSKIRICWNTADGEGGAVDSNLNQDEIP-PETQSCSL---MCPNECVMSWGL	994
32	511	SCGMGTRSRRE-RYVKQFPEDG-----SMC---KVPTETEKEKIVNEECSPSSCLVTWGE	561

Qy	995	WSKCPOSQDPHTMQRTRHL-LRPSLNSRTCAEDSQVQPCLLNENCFQFYNL---TEWS	1050
Db	562	WDECSASCGT-GMKRRHRMIKMTPADGSMCKAETTEAEKMMPE-CHTIPCLLSPWSEWS	619
Qy	1051	TCOLSENAPCGQGVRT--RLLSCVCSGDKPVSMDQCEQHNLKFPQRMSPCLVECVVNCQ	1108
Db	620	DCSVT---CGKGMRTQRMLKSAELG-----DCNE-ELEQAEKMLP---ECPIDCE	665
Qy	1109	LSGWTAWTECSTCGHGRMSRTRFIIMPTQGEGRPCPTTELTOEKTCPVTPCYSVWVLGNW	1168
Db	666	LTEWSQWSECNSTCGK-GHMIRTRMIKIEPPQFGGTACP-ETVQRTKCRVRKCLR-----	717
Qy	1169	SACKLEGGDCGEGVQIRSLSCMVHSGSISHAAGRVEDALCGEMPFDOSILKOLCSVPCPG	1228
Db	718	-----GPGWEKR-----RWKEA--REKRSSEQAKNIDNEQYP-748	
Qy	1229	DCHLTWSEWSTCELTCDIG--RSFTVGRGRSRRTFIIQSPENQDSQPQVLETRPCTG	1286
Db	749	VCLRPWTAWTECSTLCGGGIQERYVMVKRSKSTQF-----TSCKDK-KELRACNV	799
Qy	1287	GKC	1289
Db	800	HPC	802

Search completed: February 24, 2004, 01:09:33  
Job time : 64.7866 secs



D6 206 RRRGKNKERKDRSGVKD--PEARELTKKKRNRRQNRQENKYWDIQIGYQTRVVMCIN 263

1223 SVPCGCHLTENSEWSTCBLCIDGRSFTVGRQSRSTFIQSPENQDSQOVLTR 1282  
1239 SQPCGCVLKWSSWSLQCTVNGEDLGGGTQVRSRVIIQELNQLCEQMLETK 1398  
1283 PCTGGKCVHTWKASLNNNERTVWCORSQVNVGCVSPQAPAAIROCI PACRKPFYS 1342  
1299 SCYDQCYEYKXWASANKGSRTVWCORSQVNVGCVSPQAPAAIROCI PACRKPFYS 1458  
1343 CTQGGVCGCKGYTEIMKNSGFLDYC-----MKVPGSEDKKADVKNLSGKRRPVNSKIH 1397  
1459 CSEKTCHEBGEYEVMSNSSTLEQCTLIPIVWVLTMTMEDKRGDVKT-SRAVHPTQPSNP 1517  
1398 1PKG--KSLQPLDGDGKVKIYVGVSGAFILMTFLFTSYLVCKKP-KPHOSTPPQOKP 1454  
1518 AGRTWFLQPGFGDGLKTVGVGAAGFVLLIFIVSMYILACKPKPKQRRONRLKP 1577  
1455 LTLAYDGDMD 1465  
1578 LTLAYDGDMD 1588

RESULT 2  
US-10-074-566-2  
; Sequence 2, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shinkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1588  
; TYPE: PRT  
; ORGANISM: human  
US-10-074-566-2

Query Match 52.1%; Score 4296; DB 15; Length 1588;  
Best Local Similarity 50.7%; Pred. No. 0;  
Matches 766; Conservative 238; Mismatches 449; Indels 58; Gaps 20;

168 SDQNALSLCLQDSPLTVQSCIMPKDQCTSQWSSWSPCKTCRSGSLLPQFRSRNV 227  
264 KTGKAADLSFCQCKLPMTFQSVITKEQVSEWSEWSPCKTCHDWPAGTRVTRTI 323  
228 KHAIAGGKCEPPELLEKEACIVGEUULLQOCPRYSWTSEWKECONSLLEBOOPHHVVG 287  
324 RQPIGSEKCEPEFEKPELCSQDGVVPCATYGMWTTTECRVDPLLSQDQKRGNOT 383  
288 PVCGGGIQTREVTCAOS-----VPAALAAKRAKESRVERKALCVGAPLPSQLCNIPCSID 343  
384 ALCGGGIQTREVTCAQANENLSQLSTHKNKEASKPMDLKLCGTGPIPTNTQLCHIPCPT 443  
344 CIVSSWSAWGLCIHENCHBQGGKGRTRORHVLMESTGPA-----GHCPLHVSVPCEDEPN 400  
444 CEVSPWAWGPTCYENCNDQGGKGGKPKRRITNEPTGGSGVGTGNCPLHLEAIPCEEP 503  
401 CYRWLASE-GICPPDHGK-CGLGHRILKAVCQNDRGEDVSGSLC-----PVPPPERKSC 453  
504 CYDWKAVLGDCEPDNGKEGPGTQVQEVVVCINSDEEVDRLCRDAIPFIP-----VAC 558  
454 EIPCRMDCVLSEWTEWSSCSQSCSNKNSDKQTRSTILALAG-EGKPCPPSQALQEH 512  
559 DAPCPKCVLSTWSTWSSCSHTSGKTTEGQIRARSILAYAGEEGGIRCPNSSLQEV 618  
513 LCNDSHSCWQLHWETSPWPCSEDTLVTALNATIGWGEATCGVGIOTRRVFCVKSHVGOV 572  
619 SCNEHPTVTHQGTGPGQCIETSVSFNTTTWGEASCVGMQTRKVICVRNVGQV 678  
573 MTRKCPSTRTETVPFCPLPKCKDCIVTAPSEWTPCPRMCAQNAV-KOSRYRIIIOEA 631  
679 GPXKCPESLAPETVPCLLPCKKDCIVTAPSEWTPCPRMCAQNAV-KOSRYRIIIOEA 738  
632 ANGOECEDPTLYERECEDESVLCPVHWKQKWSPCILVPESVWQGITSSSEACGKLOT 691  
739 ANGRDCTDPLYEKACBAQACOSYKWKTHRRCCQLVPSVQDQSPGAGCGGSGROA 798  
692 RAVSCISDDNRSAEMELCKQTNGMPLLVQECTVPCREDCTFTAWSKFTPCSTNCATKS 751  
799 RAITCRQDGGQAGIHECLQVAGVPALTAQCIQPCQDDCQLTSWSKFSNCGDCGAVRT 858  
752 RRROLTGSKRKEKQSDILYVETELCPCEDEFISQPYGNWSDCILPEGERPHGLVR 811  
859 RRTVLGSKSKKCKKSHLYPLETQYCPCKYNAQVGVNWSDCILPEGKVLVLLGMKV 918  
812 QADSKECEGLRFRVAVACSKNGRPVDPFSCSSGYIQEKCVPICPPFDCKLSWSSWGC 871  
919 QGDIKECGQGYRYQMACYDQNGELVETSRCSHGYIEACILPCPSDCKLSWSSWGC 978  
872 SSSCGIGVIRSKWIKKPYNGRPPCKLDLKN--QVHEAVPCVSECNQYQSVVVEHWSSC 929  
979 SKSCGSGVYKRSKWLREKPYNGRPPCKLDLKN--QVHEAVPCVSECNQYQSVVVEHWSSC 1038  
930 KINNELSLRCGGGTQGRKRCV-NTADGEGAVDSMLCQDEIPPTQSCSLMCEVNC 988  
1039 KVTVMNRENGEGVQTRKVKCMQNTADGSEHVEDYLCPEMPLGSRVCKLPCPEDCV 1098  
989 MSEWGLMSKCPQSCDPHMQTRHLLR-PSLNERTCAEDSQVQPCLLNENCQFQYNLT 1047  
1099 TSEWGPWTQCVLPNQSSFRQSDPFRQPADGRCSPNAVEKPCNLNKNKYHYDNYVT 1158  
1048 EWSICOLSENAACGGVTRLLSCVSDGKPVSDQEQHLEKPKQWSIPCLVECVVNC 1107  
1159 DWSTCQLSEKAVCGNGIITRMLDCVRSQKGVLDYCEALGLEKNQWQNTSCWCECPVNC 1218  
1108 QLSGWTAWTECSQTCGHRGMRSTRFTIIMPTQGBRCPTELTQETKCPVTPCYSWYLG 1167  
1219 QLSDSWSPSECSQTCGLTGKMRRTVTPQFGDGRPCPSLMDQSKPCVKPCVRYQYG 1278  
1168 WSAKLEGGDGBGVQIRLSLSCNHSISHAAGREVEDALCGEMP-----QDSILQLC 1222  
1279 WSPQVQACQGBGTRTNISCVSDSADDFSKVYDEEFACDIELIIDGNKNVLEESC 1338

1223 SVPCGCHLTENSEWSTCBLCIDGRSFTVGRQSRSTFIQSPENQDSQOVLTR 1282  
1239 SQPCGCVLKWSSWSLQCTVNGEDLGGGTQVRSRVIIQELNQLCEQMLETK 1398  
1283 PCTGGKCVHTWKASLNNNERTVWCORSQVNVGCVSPQAPAAIROCI PACRKPFYS 1342  
1299 SCYDQCYEYKXWASANKGSRTVWCORSQVNVGCVSPQAPAAIROCI PACRKPFYS 1458  
1343 CTQGGVCGCKGYTEIMKNSGFLDYC-----MKVPGSEDKKADVKNLSGKRRPVNSKIH 1397  
1459 CSEKTCHEBGEYEVMSNSSTLEQCTLIPIVWVLTMTMEDKRGDVKT-SRAVHPTQPSNP 1517  
1398 1PKG--KSLQPLDGDGKVKIYVGVSGAFILMTFLFTSYLVCKKP-KPHOSTPPQOKP 1454  
1518 AGRTWFLQPGFGDGLKTVGVGAAGFVLLIFIVSMYILACKPKPKQRRONRLKP 1577  
1455 LTLAYDGDMD 1465  
1578 LTLAYDGDMD 1588

RESULT 2  
US-10-074-566-2  
; Sequence 2, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shinkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1588  
; TYPE: PRT  
; ORGANISM: human  
US-10-074-566-2

Query Match 52.1%; Score 4296; DB 15; Length 1588;  
Best Local Similarity 50.7%; Pred. No. 0;  
Matches 766; Conservative 238; Mismatches 449; Indels 58; Gaps 20;

2 VECIOKLNRTVTVANICHEALQPTQACILPCPRDCVWSEFLPWSNCSKCGKCLQHR 61  
89 IACIOKNDIPAEIDICIFEFKPLLEQALIPCOQDCIVSEFSAWSECSKTCGSLQHR 148

Y 62 TRAVIAPPLFGGLQCPNLTESRACDAPISCLPGBEEYFSLKVGPMWCKRLPHLKEINPS 121  
b 149 TRHVAVPQFGSGCPNLTEFQVQ---SSPCEAEELRYSLHVGMWTCSPHRSQRQA 205  
Y 122 CRT-----VLDFNSDNERVTFK-HOSYKAHHKSWAIEIGYQTRQVSCTR 167  
b 206 RRRGNKEREKDRSGVXD--PEARLKKRNRNRQNRQENKYWDIQIGYQTRVEMCIN 263  
Y 168 SDGQWAMLSLCLQDSFPLTVGSCIMPDCETSSQWSSPCKSKTRSGSLLPFGFRSRNV 227  
b 264 KTGRAADLSFCQEKLPMTFOSVITKECVSENSEWSPCKSKTCHDMVSPAGTAVRTTI 323  
Y 228 KMAIGGKCEPPELLEKACIVBELLOQCPRYKWRSEWKECOVSLLLEQDDPHHVTG 287  
b 324 RQFPFGSEKCEPEEKEPCISQSDGVVFCATYGNRTTEWTECEKRVDPILLSQDQXRRGNQ 383  
Y 288 PFCGGGOTREYVCAQS---VPAARAAKAEVRPVEKALCVGPAPLPSCNIPGSTD 343  
b 384 ALCGGGIOTREYVQVQANENLLSQLSTHKNKASKPMDLKLCTGPIFNTTQLCHIPCTE 443  
Y 344 CIVSWSAWGLCIHENCHBPOGKGRTRORHVLWSTGPA---GHCPHLVESVPCEDPM 400  
b 444 CEVSPWSAWGPTCYENCNDQCKGFKLKRRTITNEPTGGSGVTCNCPHLLLEAIPCCEPA 503  
Y 401 CYRWLASE-GICFPDHGK-CGLGHRILKAVQNDREGEVDSGLC-----PVPPPERKSC 453  
b 504 CYDMKAVRLGDCBPNGKECGPGTQVQEVVINCINDGEVDRQLCRDAIFPFP-----VAC 558  
Y 454 EIPCMDCVLSEWTESSCSOSCKMNSDKQTRSRITLALAG-EGGKPCPPSQALQEH 512  
b 559 DAPCPKCVLSTWSSCSHTCSKTTGKQIRARSILAYAGEGGIRCPNSALQEV 618  
Y 513 LNDHSCMLHWEISPWGPCSEDTLVTALNATIGWGEATCGVGIOTRPFVCVSHVGOV 572  
b 619 SCNEHPCVYHWQTPWQCIEDTSVSFNTTWTWNGEASCSCVGMQTRKVICRVNVGOV 678  
Y 573 MTKRCPDTRBETVAPCLPCKCKCIVTAFSEWTPCPRMCOAGNATV-KOSRYRIIIOEA 631  
b 679 GPKKCPESLRBETVAPCLLPCKCKCIVTAFSEWTPCPRMCOAGNATV-KOSRYRIIIOEA 738  
Y 632 ANGGECEDPTLYEERCEBDSVLCPIYRWKPKQWSPCIIIVPESVWQGTGSEACGKLQ 691  
b 739 ANGGREDTPLYEERKACEAPACQSYRWKTHKRRRCQLVPWASQDPSGAEQCGGPRQA 798  
Y 692 RAVSCISDNRSAENMECLKQTMPLLVQECTVPCREDCTFTAKSKPTPCSTCEATKS 751  
b 799 RAITCRKQDGGQAGHECLQVAGVPALTAQCLPCQDDCLTWSKFSNCGGCGAVRT 858  
Y 752 RRRLTGSRKKEKQSDLYPLVETELCPDDEFISQPYGNMSDCLPEGRREPHRGLRV 811  
b 859 RKRTLVGSKSKKECKNSHLYPLIETQYCPCKDYNAQPVGNWSDCLPEGKVEVLGMKV 918  
Y 812 QADSKECEGLFRFRAVACDKNGRPVDPSCSSSGYIOEKVIPCPCPKLSDRWSWGSC 871  
b 919 QGDIKECGGYRYQAWCYDQNGRLVETSRNCSHGVIIEACIIPCPSCKLSEWNSRC 978  
Y 872 SSSCGIGVIRSKWLKEKPYNGRCPKDLNKN--QVHEAVPCYSECNQYSWVVERWSSC 929  
b 979 SKSCSGVNVRSKWLREKPYNGRCPKDLNKNQVYVWVPCVCHSDCNQYLWVTEPWSIC 1038  
Y 930 KINNELSLRSCGGTQSRKIRCV-NTAGEGAVDSNLKNODEIPPTQSCSLMCPNECV 988  
b 1039 KYTFVNMRENCEGVQTRKVRCMQNTADGSHVEDYLCDEPEEMPLGSRVCKLPCPEDCV 1098  
Y 989 MGEWGLWSKCPQSDPHWTRHLLR-PSLNSRTCAEDSQVQPCLLNENCFOFOYNLT 1047  
b 1099 ISEWGPWTQCVLPCHQSSFRQSRADPIRQPADEGSCFNAVEKEFCNLNKNKYHYDNYT 1158  
Y 1048 EWSTCOLSENAPCGQVTRLLSVCVSGDKGKPVMSQCEQHNLEKQPMRISPCIVCEVVC 1107  
b 1159 DWSTCOLSEKAVGNGIKTRMLDCVRSBGKSDVLYKCEALGLEKNQWNTSOWVECPVC 1218  
Y 1108 QLSGWTAWTECSQTCGHGGMRSRTPIIMPOTGEGRCPCTELTQKTCPTVPCYSWVLGN 1167

Db 1219 QLSDWSPECECSQTCGLTGKMLRRRTVTQPPFGDGRPCPSPSLMDQSKPCPVKPCYRWQYQ 1278  
QY 1168 WSACKLEGDCGEGVQIRLSLSCWVHSGSISHAAGVEDALCGEMPF-----QDSILKQLC 1222  
Db 1279 WSPCQVQEAQCCGEGTRNISCVSDGSDADDKSKVDDEFCADIELIIDGNKNMVEBSC 1338  
QY 1223 SYPCPDCHLTSEWSESTCELTICIDGRSPFETVGRQSRRTFTIIQSFENQDSCPOQVLETR 1282  
Db 1339 SQPCGDCYLKDWSSNSLQCLTCVNGEDLFGGIGVRSRPVIIQELNCHLCPQWLETK 1398  
QY 1283 PCTGGKCHYTHKASLWNNERTVWCORSDGVNVTGSGSPQARPAIROCIACRKPFSY 1342  
Db 1399 SCYDGCYCYKMAWAKWSSRTVWCORSDGVNVTGSGSPQARPAIROCIACRKPFSY 1458  
QY 1343 CTQGGVCGCEKGYTEIMKNSGFLDYC-----MKVPGSEDKADVKNLSGKRPVNSKHID 1397  
Db 1459 CSEKTKCHEEGYTEWSSNSLQCLTCVNGEDLFGGIGVRSRPVIIQELNCHLCPQWLETK 1517  
QY 1398 IFKG--WSLQPLDPPGRVKIWIYGVGGAFLLIMIFLFTSYLVCKKP-KPHOSTPPOQKP 1454  
Db 1518 AGRGRTWFLQFPQDORLKTWYVGAAGFVILLIFIVSMIYLACKPKKPPQRRNNRKP 1577  
QY 1455 LTLAYDGDLD 1465  
Db 1578 LTLAYDGDAD 1588

## RESULT 3

US-10-074-566-41  
; Sequence 41, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 41  
; LENGTH: 1588  
; TYPE: PRT  
; ORGANISM: human  
US-10-074-566-41

Query Match 52.1%; Score 4296; DB 15; Length 1588;

Best Local Similarity 50.7%; Pred. No. 0;  
Matches 766; Conservative 238; Mismatches 449; Indels 58; Gaps 20;

2 VRCIQKLNRTVANEICEHFALQPPTEQAQLIPCRDCCVSEFLPWSNCSKCGKQLQHR 61  
89 IACIQKNDIPAEIDICEFEPKLEQAQLIPQDCCIVSEFSAWSECKTCGSGLQHR 148  
62 TRAVIAPPLFGGLQCPNLTESRACDAPISCIPEGEYTFSLKVGWMSKRLPHLKEINPS 121  
149 TRHVAPPQFGSGCENLTFQVCQ---SSPCEAEELRYSLHVGWSTCSMPHSQRQA 205  
122 GRT-----VLDFNSDSNERVTFK-HOSYKAHHHKSWAIEIGYQTRQVSCTR 167  
206 RRGKXKEREKDRSKGVKD--PEARLIKKBNNRQNRQENKYWDIQIGYQTRVWCIN 263  
168 SDQONAMLSLCLQDSFPLTVQSCIMPCKDCETSQWSSWSPSCSKTCRSGSLLPGRFRSRNV 227  
264 KTGKAADLSFCQCKLPMTFQSCVITKEQVSEWSEWSPSCSKTCHDMVSPAGTRVTRTI 323  
228 KMAIGGKCEPPELLBEACIVEGELLQOCPRYSWETSEWCEQVSLLEQDPHWHVIG 287  
324 RQPPISGEKEPEFEKEPCLSGDGVWPATYGVWRTTEWTCRVDPLLSQQDKRGNQT 383  
288 PVGCGGIQIREVCAQS---VPAALAAKRAKESRPRVEKALCVGPAPLPSCNIPCSSTD 343  
384 ALCGGGIQIREVYCVQANENLLSOLSTHKNKEASKPMDLKLCTGPIPTNTQLCHIPTE 443  
344 CIVSSWSANGLCIHENCHPOGKGRTRQRHVLMESTGPA---GHCPLHVESVPCEDPM 400  
444 CEVSPWSANGLCIHENCHPOGKGRTRQRHVLMESTGPA---GHCPLHVESVPCEDPM 503  
401 CYRWLASE-GICPPDHGK-CGLGHRILKAVCONDRGEDVSGSLC-----PVPPPERKSC 453  
504 CYDWKAVRLGDCPEFNGKEGCGTQVQEVVNCINSDEEVDRQLCRDAIFPIP-----VAC 558  
454 EIPCRMDVLEWTEWSSSQSOSNNKNSDKQTRRTIILALAG-EGGKPCPPSQALQHR 512  
559 DAPCPKDCVLTSTWSSCSHTCSGKTTGQIRARSILAYAGEGIRCPNSSALQEV 618  
513 LCNHDSQMLHWTSPWPCSEDTLALNATIGWNGEATCGVIGTRVFCVKSHVGV 572  
619 SCNEHPCVTVHWTGPMGQCIETISVSSFNWTTTNGEASCSVMGTRKVIQVUNVGV 678  
573 MTRKCPDSTRPVRPCLPKCKDCIVTAFSEWTPCPRMQAGNATV-KQSRRIIQA 631  
679 GPXKCPESLPETVRPCLPKCKDCIVTAFSEWTPCPRMQAGNATV-KQSRRIIQA 738  
632 ANGOCEPDLYEERECEDVSLCPVYRWKPKWSPCLLYPESVWQGITGSSACGKLT 691  
739 ANGRDCTDLYEERECEDVSLCPVYRWKPKWSPCLLYPESVWQGITGSSACGKLT 798  
692 RAVSICISDDNRSAMMECLKQTNMGMLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKS 751  
799 RAITCRQDGGQAGIHECLQYAGVPALTAQCOIPQDDCQLTWSKFSKSCNCGDCGAVT 858  
752 RRRQLTKSRKKEKQSDLYPLVETELCPDCEFIQPYGNWSDCLPEGRREPHGLRV 811  
859 KRRTLVGSKKKEKXSHLYPIETQYCPDKYNAQPVGNWSDCLPEGRREPHGLRV 918  
812 QADSEKCEGLRPRVAVACSKNGRVPDPSCSSGSGYIOEKVPCPFDCKLSWSSWGC 871  
919 QGDIKEGQGVRYQAMACYQNGRLVETSRCSHGYIEEACIIPCSDCKLSWSSWGC 978  
872 SSSCGIGVIRSKWLKBPYNGRBPCKLBN--QVHEAVPCVSECNQVSWVHWSSC 929  
979 SKSCGSGVAVSKWLREKPYNGRBPCKLBN--QVHEAVPCVSECNQVSWVHWSSC 1038  
930 KINNELSLRGGGTQSRKTRCV-NTADGGGAVDSNLNODIPEPTQSCSLMCEV 988  
1039 KVTFFNRENGCEGVQRRKRCMNTADGSEHVEDYLCDEPEMLGSRVCKLPCEDCV 1098  
989 MSEGLWSKCPQSDPHTRRTHLLR-PSLNSRTCAEDSQVQPCLLNENCFCQFQNL 1047

1099 ISEWPMPTQCVLPCNOSPFQRSGADFIQPADGRSCPNNAVEKEPCNLNKNQVHYDYNVT 1158  
1048 EWSTCOLSENAPCGQGVTRLLSCVSGDKGVPVMDQCEHNLKPEORMSPICLVECVVNC 1107  
1159 DWSTCOLSEKAVCGNGIKTEMIDCVASDGKSVLDKYCEALGLEKQWQNTSCHWVECPVNC 1218  
1108 QLSGWTATWECSTQCHGGGRSRTIIMPTQGEGRPCPTTELTOBKTCPTVPCYSWVLGN 1167  
1219 QLSWSPWSECSTQCHGGGRSRTIIMPTQGEGRPCPTTELTOBKTCPTVPCYSWVLGN 1278  
1168 WSACKLGGGCGGVQIRSLSCMVHSGSISHAAGRVEDALCGEMPF-----QDSILKQLC 1222  
1279 WSPQVCEAQCGBGTRTNISCVVSDGADDFKVVDBEFACADIELIIDGNKNWLEESC 1338  
1223 SVPCGCHTTEWSEWSTCBLTICIDORSFETVGRQSRRTFILOSFENQDSQCVLETR 1282  
1339 SQPCGDCYLKSSWSLQCLTCVNGEDLFGGIGIQRSPVIIQELNQHLCPEQMLETK 1398  
1283 PCTGKGCVHYTWKASLWNNNRTVWCQSDGVNVTGCSQAPAPAAIROICIPACRPFYS 1342  
1399 SCYDGCYFYKXWASAKGSSRTVWCQSDGVNVTGCSQAPAPAAIROICIPACRPFYS 1458  
1343 CTQGGVCGCEKGVTEIMKSNGLDYC-----MKVPGSEDKKADVKNLSGKRRPVNSKIHD 1397  
1459 CSETKCHCEGTEVMSNSTLEQCTLIPIVWVLPMTMEDKRGDKT-SRAVHPTQFSSNP 1517  
1398 IFKG--WSLOPLDPDGEVKIWWVGVSGGAEFLIMELIFTSYLVCKKP-KPHOSTPPQOKP 1454  
1518 AGRTWFLPFGFGRKLTWYGVGAAGFVLLIFIVSMYILACKPKPKPQRQNRLLKP 1577  
1455 LTLAYDGDLDL 1465  
1578 LTLAYDGDADM 1588

RESULT 4  
US-10-104-047-2928  
; Sequence 2928, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA  
; FILE REFERENCE: HI-A0105  
; CURRENT APPLICATION NUMBER: US/10104, 047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: Patent in Ver..2.1  
; SEQ ID NO 2928  
; LENGTH: 933  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-2928

Query Match 25.8%; Score 2126; DB 15; Length 933;  
Best Local Similarity 49.6%; Pred. No. 4.8e-158;  
Matches 389; Conservative 117; Mismatches 239; Indels 40; Gaps 12;

2 VRCIQKLNRTVANEICEHFALQPPTEQAQLIPCRDCCVSEFLPWSNCSKCGKQLQHR 61  
158 IACIQKNDIPAEIDICEFEPKLEQAQLIPQDCCIVSEFSAWSECKTCGSGLQHR 217  
62 TRAVIAPPLFGGLQCPNLTESRACDAPISCIPEGEYTFSLKVGWMSKRLPHLKEINPS 121  
218 TRHVAPPQFGSGCENLTFQVCQ---SSPCEAEELRYSLHVGWSTCSMPHSQRQA 274  
122 GRT-----VLDFNSDSNERVTFK-HOSYKAHHHKSWAIEIGYQTRQVSCTR 167  
275 RRGKXKEREKDRSKGVKD--PEARLIKKBNNRQNRQENKYWDIQIGYQTRVWCIN 332  
168 SDQONAMLSLCLQDSFPLTVQSCIMPCKDCETSQWSSWSPSCSKTCRSGSLLPGRFRSRNV 227

333 KTGKAADLSFCQOEKLPMTFQSCVITKEQVSEWSEWSPCSKTCHEMVSFAGTRVTRTTI 392  
228 KMAIGGKECEBELLKEACIVEGELLQOCPRYSWRTSEWKEQVSLLEQDDPHWHTG 287  
393 RQPIGSEKECEFEKEPCLSQGDGWPCATYGNWRTTEWTECRVDPILLSQQDKRGHQT 452  
288 PIVGGIGIOTREVYCAQS---VPAARALRAKEVSREVEKALCVGAPLPSOLCNIPCSTD 343  
453 ALCGGIGIOTREVYCVQANENLSQSJTHKKKASKFMDLKLCTGPIPNWTQCHIPCETE 512  
344 CIVSSWSAGLCHENCHPEPOGKGRFTRQRHVLMESTGPA---GHCPLHVESVPCEDPM 400  
513 CEVSPMSAWGPTCYENCNDQOGKKGPKLRKITNEPTGGSGVTGNCPELLEAIPCPEPA 572  
401 CYEWLASE-GICPPDHCK-CGLGHRILKXAVCONDRGEDVSGSLC-----PVPPPERKSC 453  
573 CUDKAVRUGNCEPDNGKSCGPGTQVQEVVVCINSQGEVDRQLCDAIFPIP-----VAC 627  
454 EIPCRMDCVLSEWSSQSCSNKNSDKGTRSRITLALAG-EGGKPCPPSQALQHR 512  
628 DAPCPKDCVLSWTSSCSHTCSGKTTEGKQIRARSILAYAGEGGIRCPNSSALQEV 687  
513 LCNDSHCOMLHNETSPWGCSEDTLTALNATIGWNGEATCGVIGIOTRRVFCVKSHVGOV 572  
688 SCNEHPCITYHMQTGWGQCIEDTSVSSFNITTTWNGEASCSCVGMQTRKVICRVNVGOV 747  
573 MTKRCPSDTRPVRPCFLPKCKDCIVTAFSEWTPCPRMCOAGNATV-KOSRYRIIOEA 631  
748 GPKCPESELRPETVRCLPKCKDCIVTAFSEWTPCPRMCOAGNATV-KOSRYRIIOEA 807  
632 ANGQCPCPTLVEERCEVDVSLCPVYRWKPKQWSPCLVPSVWQGITGSSEACGKGLQT 691  
808 ANGGRCCTDPLVEEKACEAFACQSQVYRWKTHKRRRCQLVPMSVQDQSPGAEQCGPGRQA 867  
692 RAVSCISDDNRSAEMECILKQINGWELVQECTVPCREDCTFTAKSKFTPCSTNCEATKS 751  
868 RAITCKQDGGQAGIHECLQYAGVPFALTOAQIFCQDDCQLTWSKSSNGDCGAVRT 927  
752 RRQL 756  
928 RRRL 932

RESULT 5  
US-09-864-761-43221  
Sequence 43221, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
SEQ ID NO 43221  
LENGTH: 247  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC011231.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1  
OTHER INFORMATION: SWISSPROT HIT: P35446, EVALUE 5.00e-10  
OTHER INFORMATION: EST\_HUMAN HIT: AW612526.1, EVALUE 3.00e-27  
US-09-864-761-43221

Query Match 10.3%; Score 852; DB 9; Length 247;  
Best local Similarity 100.0%; Pred. No. 1.1e-58;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRICOKLNRVTVANEICHFALQPPTEQACILPCPRDCVWSEFLPWSNCSKCGKGLQH 60  
DB 95 MVRICOKLNRVTVANEICHFALQPPTEQACILPCPRDCVWSEFLPWSNCSKCGKGLQH 154  
QY 61 RTRAVIAPPLFGGLQCPNLTESRACDAPISCPGEEYTFSLKVGWPSKCRPLHKEINP 120  
DB 155 RTRAVIAPPLFGGLQCPNLTESRACDAPISCPGEEYTFSLKVGWPSKCRPLHKEINP 214  
QY 121 SGRVLDVFNDSNERNVTFKHQSYPKHHHKSWSA 153  
DB 215 SGRVLDVFNDSNERNVTFKHQSYPKHHHKSWSA 247

RESULT 6  
US-09-321-987B-2  
Sequence 2, Application US/09321987B  
Patent No. US20020102210A1  
GENERAL INFORMATION:  
APPLICANT: Kimble, Judith E  
APPLICANT: Belloch, Robert H  
TITLE OF INVENTION: Agent and Method for Modulating Cell Migration  
FILE REFERENCE: 960296.95386  
CURRENT APPLICATION NUMBER: US/09/321,987B  
CURRENT FILING DATE: 1999-05-28  
PRIOR APPLICATION NUMBER: 60/087,170  
PRIOR FILING DATE: 1998-05-29  
PRIOR APPLICATION NUMBER: 60/129,023  
PRIOR FILING DATE: 1999-04-13  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 2150  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-09-321-987B-2

```
Query Match      8.4%; Score 694; DB 9; Length 2150;
Best Local Similarity 22.9%; Pred. No. 4.9e-45;
Matches 351; Conservative 133; Mismatches 556; Indels 492; Gaps 84;

QY 18 CEHFALQPTQACLIQPCR-----DCVSEFLPWSNCKGCGKKLQHRTRAVIAP-PLF 71
DB 571 CHH-----GACVRLAPESLTKIDQGWDRWSGECSTCGGVQKGLRDCDSPKRN 622
QY 72 GGLQCPNLTES-RACDAPISCPLEEYTFSLKVGPMKRLPHLKEINPSGRTVLDPNS 130
DB 623 GKKYCVGQREYRSCNTQ-ECFMDTQPY-----REVQCS-----EFN- 658
QY 131 DSNERVTFKHOSYAHHSKSWAIEIGYQTRQVSCSTRSDGQNALSL-----CLQD 181
DB 559 --NKDIGIQQVASTNTHWPRIYANVNERCKLYC-RLSGSAFYLLRDKVDTGTPCDRN 715
QY 182 SFPLTVOSCMKDC-----TSQWSSWSPCKTCRSGSLLP-----219
DB 716 GDDICVAGACMPAGCDHQLHSTLRDXGCVGGDSCKVKVGTFFNQGTFGYNEVMKIP 775
QY 220 FRSSRNKVMHAIQGGKECEPHELLEKEACIVEGELLQOCPRYSWRTSEWKECQVSLLEQ- 278
DB 776 AGSANIDIRQKGYNNKEDDNYLSLA--ANGEFLN-----GHFQVSLARQOI 822
QY 279 --QD-----PRHWHTGP-----VCGGIQTREV---YCAOSVPAALRA 313
DB 823 AFQDTVLEYSGDALIERINGTGPIRSDIYVHVLVSGSHPPDISYEWTAAPNA-----877
QY 314 KEVSRPVEKAL-----CVGPAP-----LPSQ- 334
DB 878 --VRIPISSALYLWRVTDWTPECDRAGQSQKLMCLDMSTRHQSHDRNCONVLKPKQA 935
QY 335 --LCNIPGSTCIVSSWSAWGLC-HENCHEPQGGKGFRTQRQHVLMESTGPAGCHPLVE 392
DB 936 TRMCNIDCSTWITEDVS-----SCSAKCGSGQRQRVSCVVMGEGDRTQTPASEHLD 987
QY 393 --SVPCDEPMCY-----RWLASGI-CFDPHGKGLGHRILKAVCONDRGEDVSGSLC- 442
DB 988 RNSKPSDIASCIYDCSGKMYGWTSCSETCSGNGKMR--KSYCVDDSNRRVDSLCG 1045
QY 443 -PVPPPERKSCIEPCRMDCVLSWT--EWSSCSOSCSNKSNGDKQTRSTIILAGE-- 497
DB 1046 REQEATERECNRIPC-----PRWTVGHSECSRSCDG---GVKWRHAQCLDAADRET 1095
QY 498 GKGKPPSQALQEHRLCNHDSMOLHETSPWGPCSEDTLVLTALNATIGNWGEATCGVGI 557
DB 1096 HTSRGGAQT-QEH-CNEHAC--TWQFGWSDCS-----AKCGDGV 1133
QY 558 QTRRVFCVKSIVGQWTKRCPDSTRPTETVPCFLPKKDCIIVTAFSWTFCPRMQ- 613
DB 1134 QYRDANCIDTRHSVLPBHRCLKMEKIIT-KPCH---RESCPKYKLGWSQCSVSCEDGWS 1189
QY 614 -----AGNAT-----VKOSRYRIIQAANGQBCP-----639
DB 1190 SRRVSCVSGNGTEVDMSLCGTASDRPASHQTCNLG-TCPPWRNTDMSACSVSCGIGHRR 1248
QY 640 --DTLYBERCEDVSLC-----PVYRWKPKWSPCILVPESVWQGITGS 682
DB 1249 TTECIYREQSV-DASFCQDTRWPKTSQTCHELLPCTSWKPSHWSPC-----S 1293
QY 683 EACGKGLQTRAVSC-----ISDONSRAEMWE-CLKQTNMGMLPLAVQSCVPCRE 729
DB 1294 VTCSGIQTRSVSCSTRSEGTIVDEYFCDNTRPLAKTEKTDCTDGPVLRQ- 1349
QY 730 DCTFTAMS--KFTPCSTNCEATKSRRLQTKSRKKEKCD------SDLYPLVET 777
DB 1350 DVPEIRWATGWTACATCG-----NGTQRIILKCRDHVRDLDPDECNHLDEVST 1400
QY 778 ELCPDDEFISQYGNWSDCILPEGRBPHRLRQVADSKECGELRPVAVACSDKNGRPV 837
DB 1401 RNCRLDCSYKMAEWEECPATCG--THVQOSRNVTCVSAEDGGRTILKQVDC-DVQKRP 1458
```

## RESULT 7

```
US-09-800-729-155
; Sequence 155, Application US/09800729
; Patent No. US20020068319A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 155
; LENGTH: 2165
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-155
```

```
Query Match      8.4%; Score 694; DB 9; Length 2165;
Best Local Similarity 22.9%; Pred. No. 5e-45;
Matches 351; Conservative 133; Mismatches 556; Indels 492; Gaps 84;

QY 18 CEHFALQPTQACLIQPCR-----DCVSEFLPWSNCKGCGKKLQHRTRAVIAP-PLF 71
DB 586 CHH-----GACVRLAPESLTKIDQGWDRWSGECSTCGGVQKGLRDCDSPKRN 637
QY 72 GGLQCPNLTES-RACDAPISCPLEEYTFSLKVGPMKRLPHLKEINPSGRTVLDPNS 130
DB 638 GKKYCVGQREYRSCNTQ-ECFMDTQPY-----REVQCS-----EFN- 673
QY 131 DSNERVTFKHOSYAHHSKSWAIEIGYQTRQVSCSTRSDGQNALSL-----CLQD 181
```

674 --NKDIGQVASTNTHWPYKAVNAPNERCKLYC-RLSGSAAFYLLRDKVVDGTPCDRN 730  
182 SPLTLVQSCIMPCKCE-----TSOWSSWSPCSCTCRSGSLLP----- 219  
731 GDDICVACAGPAGCDHQLHSLTRDKCGVGGDDSSCKVVGTFNEOGTGYNEVMKIP 790  
220 FRSRNRVKNHAIQGGKCPPELLEKACIVGELLQOCPRYSWRTSEWKEQCVSLILLEQ- 278  
791 AGSANIDIRQKYNMKEDDNYLSRA--ANGEFLN-----GHFQVSLARQOI 837  
279 --QD-----PHWHVTGP-----VCGGGIQTREV---YCAQSVPAALRA 313  
838 AFQDTVLSYSGSDAILIERINGTGPTRSDIYVHVLVSGSHPPDISYEWTAAPNA----- 892  
314 KEVSPVSKAL-----CVGPAP-----LPSQ- 334  
893 --VIRPISALYLWRTDTWTECDRACRQGOQSKMLCLMDSTHRQSHDRNCQVLPKQA 950  
335 --LCNIPGSTCIVSSWSAWGLCIHENCHEPQKKGTRQORHVLMESTGPAGCHPLVE 392  
951 TRMCMIDCSTRWITDVS-----SCSAKCGSQKQRQVSCVMEGDROTTPASEHLCD 1002  
393 --SVPCDEPMCY-----RWLASEGI-CFPDHGKGLGHLIRLKAQVQNDRGEDVSGSLC- 442  
1003 RNSKPSDIASCVIDCSGRKNVGEWTSCTCGSNGKHE--KSYCVDDSNRRVDESICG 1060  
443 --PVPPPPERKSCIEPCRDVCLSEWT--EWSGCSQSCSNKNSDKQTRSRITILAGE-- 497  
1061 REQKEATRECNRIEPC-----PRVYGHWSECSRCDG-----GYQRHAQCLDAADRET 1110  
498 GKGPFPQQAQLEHRLCNDHSCMQLHWETSPWQCSIEDTLVATLNATIGWNGEATCGVI 557  
1111 HTSRGPAQT-QEH-CNEHAC--TWQFGVWSDCS-----AKCGDGV 1148  
558 QTRRVFCVSHVQVQWTKRCDPSTRPETVRPCFLPKKDCIVTAPSEWTPCPRMCQ----- 613  
1149 QYRDANCNTDRSRVLPFRCLKMKEXIIT-KPCB--RESCKYKLGWSQCSVSCDGS 1204  
614 --AGNAT-----VKOSRYRIL1QEAANGQCEP----- 639  
1205 SRRVSCVSGNGTEVDMSCLGITASDPASHQTCNLG-TCPFWRNTDMACSVCSGIGHRR 1263  
640 --DTLYERECDEVSIC-----PVYWRKPKQWSPCILLVPESWQGITSS 682  
1264 TTECIYREQSV-DASFCGDTKMPETSQTCHELLPCTSWKPSHWSFC-----S 1308  
683 EACGKGLQTRAVSC-----ISDDNRSAMMB-CLKQTNGLPLLVQECTVPCRE 729  
1309 VTCGSGIQTRSVSCTRGSEGTIVDEYFCDRNTRPRLKTKCTCKDTCGPRVLQK-----LQA 1364  
730 DCTFTAWS--KFTPTCSINCEATKSRRLQITGSRKKEKQD-----SDLYPLVET 777  
1365 DVPPIRWATGPWTACATCG-----NGTQRELLKCRDHVRDLPDEYCNHLDKEVST 1415  
778 ELPCDFEISQPYGNWSCILLPEGRBRPHRGLRVQADSKGEGELRFAVACSKNGRPV 837  
1416 RNCRLDCSYWKNMAEWECAPATCG-THVQOSRNVTCVSAEDGGRTILKDVDC-DVQKRPT 1473  
838 DPFCSGSGYIQEKVIFCP-FDCKLSW--SSWGSQSSCGIGVIRSRKWLKPKYNGG 894  
1474 SARNCR-----LEPCPKGEHIGSWITGDSKCSACSGGWRRRSVCTSSSCDET 1524  
895 RCPCKLDLKNQVHEAVPCYBCNQ-----YSWVEHWSCKINNELSLRSCGGTQS 946  
1525 RK-PRV-----FDKNEBLCPPLTNNSWQISPWTHC-----SVSCGGGVQR 1564  
947 RKIRCVNTADGEGAVDSNLCNDBEI-----PRETQSC-----SL 981  
1565 RKINWEDVLSGR-----KQDDIESEIKPREQRCDEMPPCRSHVHNKTSASMTSL 1615  
982 MCPNECVMS-----EW--GLWSKCPQSCDPHTWQRTRHLLPSLN-----S 1021  
1616 SSSNNTSSASASSLPILPVVWSQTSAMWASAKCGRGT-KRRVVECVNPSLNVIVAS 1674

QY 1022 RTCAEDSOVQPC-----LLNENCFOQFOYNLTJNSTCOLSENAPCGQGVTRTELLSCVSCDGK 1077  
Db 1675 TEC--DOTKXPEVEVRCRTHKC--PAWKTTTWTSSSCSVT-----CGRIRREVOCTRGKRN 1726  
QY 1078 PYSMDQCEQHNLEKPMQMSIPCLVECV--VNCQLSGW--TAMTECSQTCGHGGRMSRTFI 1134  
Db 1727 LVSDSECN-----PKTKLNSVANGCFPVACPAYRWNVTPWSKCKDEACARGQKQTRRVHC 1779  
QY 1135 IMPQTGEGRCPTELTQEXT-----CPVTPC-YSWVLGNWSACKLEGGDCGEGVQTRSLSC 1189  
Db 1780 ISTSGRAAPRMCELARAPTSIRECDTSNCPYEWVPGDMQTC-----SKSCGEGVQTRVRC 1836  
QY 1190 --MVHSGSISHAAGRYVEDALCGEMPQDSILKQLCSV-PCPGD-----CHLT-----EWS- 1236  
Db 1837 RRKINFNSTIPIFMLED-----EPAVPEKCELPKPNESQTCNELNCPDSEFFKNF 1888  
QY 1237 -EWSCTELTC-----IDGRSPETV 1254  
Db 1889 GPMGECCKNCGQGIARRRVKCVANDGRRVERV 1920

RESULT 8  
US-10-274-639-10  
; Sequence 10, Application US/10274639  
; Publication No. US20030232349A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: DELEGEANE, Angelo M.; GANDHI, Ameena R.  
; APPLICANT: HAFALIA, April J.A.; LU, Dylung Aina M.  
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.  
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.  
; APPLICANT: NGUYEN, Daniel B.; LEE, Ernestine A.  
; APPLICANT: KHAN, Farrah A.; YUE, Henry  
; APPLICANT: AU-YOUNG, Jennifer K.; GRIFFIN, Jennifer A.  
; APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi  
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha  
; APPLICANT: DING, Li; KEARNEY, Liam  
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.  
; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.  
; APPLICANT: BURFORD, Neil; WALIA, Narinder K.  
; APPLICANT: LAL, Preeti G.; LEE, Sally  
; APPLICANT: TODD, Stephen; LO, Terence P.  
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.  
; APPLICANT: AZIMZAI, Valda; JU, Yan  
; TITLE OF INVENTION: PROTEASES  
; FILE REFERENCE: PI-0167 USA  
; CURRENT APPLICATION NUMBER: US/10/274,639  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: PCT/US01/22397  
; PRIOR FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: US 60/220,063  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 60/221,680  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: US 60/223,544  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/224,717  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: US 60/225,988  
; PRIOR FILING DATE: 2000-08-16  
; PRIOR APPLICATION NUMBER: US 60/227,568  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PERL Program  
; SEQ ID NO 10  
; LENGTH: 1916  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030232349A1 2994162CDI  
US-10-274-639-10

Query Match		8.2%; Score 674; DB 15; Length 1916;
Best Local Similarity		21.4%; Pred. No. 1.9e-43;
Matches	290; Conservative	123; Mismatches 419; Indels 526; Gaps 70;
QY	201 WSS---	WSPCKTRSGSLLPGRSRNRNVKHAIGGKCEPPELLEKACIVEGELL---254
Db	862 WNSHGPWQACSKPCQ---	GERKR-----KLVCTRESDLTVS 895
QY	255 -QCCPRYSWRTSEWKEQVSLLEEQDPHWHVT-----	GPVCGGIGTREVYCAQSVPA 308
Db	896 DQCDRLPQGHITPCGTDC-----	DLRHVASRSECSAQCGLYRTLDIYCAK-----945
QY	309 AALRAKEVSRVEKALCVGPAPLPSCNLTPCSDTIVSSW--	SANGLCIHENCHEPQK 366
Db	946 -YSRLDGKTEKVDGFC--SSHPKPSN--REKSGECNTGGWRYSAWTEC-----	990
QY	367 KGFRTQRHVLMESTGPAGHCPHLVESVPCEDPMCYRWLASGICFPDHGKGLGHRILK 426	
Db	991 -----	SKSCDGTQRR 1002
QY	427 AVCQNDRGEDVSGSLCPVPPPERKSC-EIPCRMDCVLSEWT--	EWSSCSQSCSNKNSDG 483
Db	1003 AICVNTENDVLDDSKTHQEKVTTQRCSEFFC-----	PQWKGDMSECLVTC-----G 1050
QY	484 KQTESRTILALAGE---GGKPCPSQALQEHRLNDHSCMQLHETSMPGCPSEDTLVTA 540	
Db	1051 KGHKRWQVQFGEINLRNDRMCDDETPTSMQTCQPECAS--	WQAGPMGQCS-----1101
QY	541 LNATIGNWGEATCGVIGTTRVFC-VKSHVQVYMKRCPDSTRPETVRPCFLPCKKDC--	597
Db	1102 -----VTCQGGYQLRAVKCIIGTVMSVDDNDCAATRTPTDQCELP--	SCHP 1148
QY	598 -----IVTAFSEWTPCPRCVQAGNATVKSRYRILIOEAANGQGCPT 641	
Db	1149 PPAAPETRSTYSAPRTQWRFGSTPCSATCGKG-----	TRMYVSCRDEG-----S 1196
QY	642 LYBEREC-----EDVSLCPVYRWKPKQWSPCILVPESVWQGTGSEACGKGLQTR 692	
Db	1197 VADSSACATLPVPYAKBECSTVPCGQWALDWSC-----	SVTCQGRATR 1242
QY	693 AVSCISDNRSAENMECLKQTNMGMLLVQSTVPCREDCTFTAWSPTPCSTNCEATKR 752	
Db	1243 QVMCVNTSDHVIDRSEC--DQDYIPKTDQCSM-----	SPC-----1276
QY	753 RRLTGSRKKEKQDSDL--YPLVETELCPDEFISQPV-----	GNWSDCILPEG 801
Db	1277 -----PQTFDGLAQHPQONEDYPRGASPSRTHVLGNQWRTPWQAC-----	1321
QY	802 RREPHRLRVOADSKCEGGLRFRVACSKNG-----	RPVDPFSCSSGYIQEX 851
Db	1322 -----SSTCAGGSRVRVVQDENGVTANDCVRIKPDQACESG-----	1362
QY	852 CVIPCPDCKLSDWS--SWGSSSSCGIVRIRSKWLKXPYNGRPPCKPLDLKNQVHEA 909	
Db	1363 ---PCP-----QWYGNWGECKLQCGGIRTR--	LVCQRNGERFP--DUSCSILDK 1408
QY	910 VPCYSECNQY-----SWVYEHWSCKINNELSLRGGGTQGRKIRCVNTADGEGGAVD 963	
Db	1409 PPDEQCNTACPHDAWSTGPMWSSC-----	SVSCGRGHQNRVYCMK-----DGSHLE 1458
QY	964 SNLCNQBEPETQSC-SLMCPNCEVSEW--GLWSKCPQSCDPHTNQRTHLLRPSLN 1020	
Db	1459 SDYCKHLAKPHGRKCRGRCP-----	KWAGAWSQCSVSCRGVQQRH-----VQCQIG 1508
QY	1021 SRTCAEDSQVQCLLNEN-----CFQFOYNLTWSTQCLSENAPCGQGVTRLLSCV 1072	
Db	1509 THKTARETCNPTRPSEBRDQCPRCFLYTWRAEWQEC-----	TKTCGGSRYRVKVCV 1564
QY	1073 -----CSDGK--PVSMQDCEQHNLEKFORMSIPCLVECVNQLSGWTA--	WTEC 1118
Db	1565 DDNKNEVHGARCDSKRPVDRSECSLQ-----	PC-EYV-----WITGEWSEC 1605
RESULT 9		
US-09-938-330-25		
; Sequence 25, Application US/09938330		
; Patent No. US20020115838A1		
; GENERAL INFORMATION:		
; APPLICANT: Walke, D. Wade		
; APPLICANT: Hilbun, Erin		
; APPLICANT: Scoville, John		
; APPLICANT: Friedle, Carl Johan		
; APPLICANT: Hu, Yi		
; APPLICANT: Turner, C. Alexander Jr.		
; TITLE OF INVENTION: No. US20020115838A1el Human Proteases and Polynucleotides Encodin		
; FILE REFERENCE: LEX-0237-USA		
; CURRENT APPLICATION NUMBER: US/09/938,330		
; CURRENT FILING DATE: 2001-08-22		
; PRIOR FILING DATE: 2000-08-22		
; PRIOR APPLICATION NUMBER: US 60/227,104		
; PRIOR FILING DATE: 2000-08-22		
; PRIOR APPLICATION NUMBER: US 60/233,796		
; PRIOR FILING DATE: 2000-09-19		
; NUMBER OF SEQ ID NOS: 26		
; SOFTWARE: FastSeq for Windows Version 4.0		
; SEQ ID NO 25		
; LENGTH: 1907		
; TYPE: PRT		
; ORGANISM: homo sapiens		
US-09-938-330-25		
Query Match	8.2%; Score 673; DB 9; Length 1907;	
Best Local Similarity	21.4%; Pred. No. 1.9e-43;	
Matches	290; Conservative	123; Mismatches 419; Indels 526; Gaps 70;
QY	201 WSS---	WSPCKTRSGSLLPGRSRNRNVKHAIGGKCEPPELLEKACIVEGELL---254
Db	853 WNSHGPWQACSKPCQ---	GERKR-----KLVCTRESDLTVS 886
QY	255 -QCCPRYSWRTSEWKEQVSLLEEQDPHWHVT-----	GPVCGGIGTREVYCAQSVPA 308
Db	887 DQCDRLPQGHITPCGTDC-----	DLRHVASRSECSAQCGLYRTLDIYCAK-----936
QY	309 AALRAKEVSRVEKALCVGPAPLPSCNLTPCSDTIVSSW--	SANGLCIHENCHEPQK 366
Db	937 -YSRLDGKTEKVDGFC--SSHPKPSN--REKSGECNTGGWRYSAWTEC-----	981
QY	367 KGFRTQRHVLMESTGPAGHCPHLVESVPCEDPMCYRWLASGICFPDHGKGLGHRILK 426	
Db	982 -----	SKSCDGTQRR 993
QY	427 AVCQNDRGEDVSGSLCPVPPPERKSC-EIPCRMDCVLSEWT--	EWSSCSQSCSNKNSDG 483



b 994 AICVNRNDVLDSDCKTHQKVTIQRCEFFPC-----POWKSQDMSCLVTC-----G 1041  
y 484 KOTRSRTILALAGE---GGKPCPPSQALQEHRLCNDHSCMQLHWETSPWGPCSEDTLVTA 540  
b 1042 KGHKRWQVWQFGEDELNDRCMDPETKTSMTQCPKAS--WQAGPWQCS----- 1092  
y 541 LNATIGWNGEATCGVIGIOTRRVFC-VKSHVGQVMTKRCPDSTRPVRPCLPKCKDC-- 597  
b 1093 -----VTCGGQYQLRAVKCIIGTVMYVDDNDCAATRPDTQDCCLP---SCHP 1139  
y 598 -----IVTARSEWTPCPRMCOAGNATVKQSRVRIIIQEAANGQCEPDT 641  
b 1140 PPAAPETRSTYSAPRTQWFGSWTPCSATCKG-----TMRVSVCRDENG-----S 1187  
y 642 LYBEREC-----EDVSLCPVYRWKPCQWSPCILVPESVMOGITSSACGKGLQTR 692  
b 1188 VADESACATLPRPAKECSVTPCGWKALDWSSC-----SVTCGGQGRATR 1233  
y 693 AVSCISDDNRSAMMECLKQTMGMPLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSR 752  
b 1234 QWCVNYSDHVIDRSEC--DQYIIPETDQCSM-----SPC----- 1267  
y 753 RRLATGKRKEKQCDSDL--YPLVETELPCDEFISQPY-----GNWSCILPEG 801  
b 1268 -----PORTDPSGLAHPFQNEIDYRPSASPSRTHVLGGNQWRTGPGAC----- 1312  
y 802 RREPHRGLRVQADSKEGELRFRVAVASDKNG-----RPVDPFCSGSGYQEK 851  
b 1313 -----SSTCAGGSQRRVVVQDENGTYTANDCVERIKPDEQACESG----- 1353  
y 852 CVIPCFDCKLSDWS--SWGSCSSCGIGVIRSKWLKEKPYNGRCPKLDLKNVHEA 909  
b 1354 ---PCP-----QWAGNWGECKLCCGGIRTR---LVVQSRNGERFP--DLSCILDK 1399  
y 910 VPCYSECNOY-----SWVEHWSSCKINNELSLRCGGGTQSRKIRCVNTADGEGAVD 963  
b 1400 PPDREQCNTHACPHDAANSTGPWSSC-----SVSCGRGHQKRVNYCMWAK---DGSHE 1449  
y 964 SNLQNQDEIPPETQSC-SLMCPNECVMSW--GLWSKCPQSCDPTHQWQRTHLRLSLN 1020  
b 1450 SDYCKHLAKPHGRKCRGRCP-----KWAGAWSQCSVSGRGVQORH---VGCQIG 1499  
y 1021 SRTCAEDSQVOPCLLNN-----CFQOYNLTWSTCOLSENAPQGGQVRRLLSCV 1072  
b 1500 THKIARETECNPYTPRESERDCQGRCPPLYTWRAEWQEC---TKTCGEGSRVYKVVCV 1555  
y 1073 -----CSDGK-PVSDMOCEQHNLERPORMSIFCLVECVNQLSGWTA--WTEC 1118  
b 1556 DNKNVEHGARDVSKRPVDRSCSLQ-----PC--EYV-----WITGEWSEC 1596  
y 1119 SOTCGHGRMSRTRFIIMPTQEGRCPTELTQEKI--CP-----VTPCY-----S 1162  
b 1597 SVTCGKYQKRLVSCSEIYTKEN---YESYQTTINCPTQTPSPVHPCVLRCPVSAT 1652  
y 1163 WVLGNWSACKLEGGCGGVQIRLSUSCMVHSGSISHAA----- 1200  
b 1653 MRVGNWGSV---SCGVGMQORSVQCLTNEQPSHLCHTLKPEERKTCRNVCNCLPQ 1709  
y 1201 -----GRVEDALCE--MPQDSTILKQICSVPCPDCHLTWSEWSFELTCTIDGR 1249  
b 1710 NCKEVRLKLGASED---GEYFLMIGKLLKIFCA-----GMHSDHPKEYVT--LVHGDSE 1759  
y 1250 SP-ETVGRQSRRTFIISFENQDSC-----POQVLET----- 1281  
b 1760 NPSEYVGHRLHNTEPCYPNGSRDRDCQCKRDYTAAGFSSFKIRIDLTSMQIITDQLQA 1819  
y 1282 ----RP---CTGGKCYHT-----WKASLWN-----NEETVWCQS-----D 1312  
b 1820 RTSEGHVPFFATAGCYGAACKPOGRFSINLYGTGLSLTESARMISQNAVSDIKKSPD 1879  
y 1313 GVNVTGGSPQARPAARICQIPACRKPESYCTQGGVCG 1350

Db 1880 GTRVVGKC-----GGYCG 1892

## RESULT 10

US-09-918-171A-13  
; Sequence 13, Application US/09918171A  
; Patent No. US20020110894A1  
; GENERAL INFORMATION:  
; APPLICANT: Apte, Suneel  
; APPLICANT: Hurskainen, Tiina L.  
; APPLICANT: Hirohata, Satoshi  
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases  
; FILE REFERENCE: 26473/04193  
; CURRENT APPLICATION NUMBER: US/09/918,171A  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: 09/369,364  
; PRIOR FILING DATE: 1999-08-06  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1882  
; TYPE: PRT  
; ORGANISM: Homo sapiens ADAMTS-9  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (468)  
; OTHER INFORMATION: Xaa = Cys  
; NAME/KEY: MOD RES  
; LOCATION: (521)  
; OTHER INFORMATION: Xaa = Tyr  
US-09-918-171A-13

Query Match 8.0%; Score 657; DB 9; Length 1882;

Best Local Similarity 21.2%; Pred. No. 3.3e-42;

Matches 288; Conservative 123; Mismatches 421; Indels 526; Gaps 70;

QY 201 WSS---WSPCKTCSRGSLLPGFRSRNRVKNEMAGGKCEPPELLEKEACIVEGELL--- 254  
Db 828 WNSHGWQACSPQ-----GERKE-----KLVCTRESQJTVS 861  
QY 255 -QCPRYSNRTSEWKECQVSLLEOODPHMHTV-----GPGCGGIQTRVYCAQSVPA 308  
Db 862 DQRCDLRQPGHITPCGTGC-----DLRWVARSSECSAQGLGYRTLDIYCAK----- 911  
QY 309 AALRAKEVSRPEKALCVGPAPLPSOLCNIPGSTCIVSSW--SANGLCIHENCHEPQK 366  
Db 912 -YSLDGKTEKVDGFC--SSHKPSN--REKSGECNCGMYSAWTEC----- 956  
QY 367 KGFRTQRHVLMESTGPGAGHCPHVESVPCEDPMCYRWLASEGICFPDHGKGLGHRILK 426  
Db 957 -----SKSCDGGTQRRR 968  
QY 427 AVCONDREDEVSGLCPVPPPERKSC-EIPCRMDCVLSEWT--FWSSCSQSCSNKNSDG 483  
Db 969 AICVNRNDVLDSDCKTHQKVTIQRCEFFPC-----POWKSQDMSCLVTC-----G 1016  
QY 484 KOTRSRTILALAGE---GGKPCPPSQALQEHRLCNDHSCMQLHWETSPWGPCSEDTLVTA 540  
Db 1017 KGHKSQVWQFGEDELNDRCMDPETKTSMTQCPKAS--WQAGPWQCS----- 1067  
QY 541 LNATIGWNGEATCGVIGIOTRRVFC-VKSHVGQVMTKRCPDSTRPVRPCLPKCKDC-- 597  
Db 1068 -----VTCGGQYQLRAVKCIIGTVMYVDDNDCAATRPDTQDCCLP---SCHP 1114  
QY 598 -----IVTARSEWTPCPRMCOAGNATVKQSRVRIIIQEAANGQCEPDT 641  
Db 1115 PPAAPETRSTYSAPRTQWFGSWTPCSATCKG-----TMRVSVCRDENG-----S 1162  
QY 642 LYBEREC-----EDVSLCPVYRWKPCQWSPCILVPESVMOGITSSACGKGLQTR 692  
Db 1163 VADESACATLPRPAKECSVTPCGWKALDWSSC-----SVTCGGQGRATR 1208  
QY 693 AVSCISDDNRSAMMECLKQTMGMPLLVQECTVPCREDCTFTAMSKFTPCSTNCEATKSR 752

```

; TYPE: PRT
; ORGANISM: H. sapiens
Us-10-213-509-5

Query Match      8 0%; Score 655.5; DB 14; Length 4123;
Best Local Similarity 21.9%; Pred. No. 1.2e-41;
Matches 390; Conservative 149; Mismatches 567; Indels 677; Gaps 101;

QY 28 EQACLLPCPRDCVVSFPLPWSNCSKCGKKLQHRTRAVIAPPILFGGLQCP-NTLESRACD 86
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2507 EDGCV-----DCVLAFWVMWSCSRSCGLGLTFQRELLRPPLPGG-SCPRDRFRSQCF 2560
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 87 APISCPGLBEEYTFSLKVGPSKCRPLPHLKEINPSGRTVLDFNSDSENERVTFKHQSYK 146
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2561 VQ-ACV-----AGAWAWAAGPCSVSCGG-----2585
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 147 HHSKSWAIBIGYQTVQSVCTSRDQGN--AMLSLCLQDSFPLTVQSGIMPDKCETSOWS-- 202
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2586 -----GHQSRQSRSCVDPPPKNGGAPCPGASQERAPCGLQPCSGGKGVLGWGHG 2635
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 203 -----SW-----SPCSKTCRS-----GSLLPGRFSRSRNVK 228
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2636 GSTVGTGRGLPAPRLTWCPSTRLRAGPCVCRSPVEGAGAPMTLLPG--SQGQKL 2693
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 229 HMAIGGG-----KCPCLLEKE-----245
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2694 QWALCSLSLLCPLGLSALFHLPGCRCPGGLLLHDTCLPLSEPCLVGBELKWPVS 2753
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 246 -----ACTIVE-GELOQ--CP-RYSWRT-SEWKECQVSLLEQQDPHHVHTGPVCGG 292
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2754 FLLGNCQCVCEKGEILLCPGGCFLPCGMSAWSSWAPCDRS-----CGS 2797
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 GICTREYVCAQSVPAALRAK-EVSRP-----VERALCVGPAPLPSQLCNIP 339
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2798 GVRAR--FRSPNPPAANGAGPCEGRDQELQGHCTVCGTGIAGSLGAGVPPSSQFTCLR 2855
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 340 C-----STDCIV-----SSWAWGLCIHENCHEPQKKGFTRQHVLMESTGPAGHC 387
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2856 THGWGPTDHTSWGIEVFGWTPMTWSSC--SQSLAPGGGPGWRSR--LCPSPGDSS-C 2911
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 P-HLVSVPCEDPM-CYRWLASGICFPD-----HGK-----417
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2912 PGDATQEEPCSPPIETGFCAPGCTCPGLFLHNASCLPRSQPCQLHGQLYASGAMARL 2971
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 418 -----CGLGHRILKAVCONDREGEDVSGSLCPVPPPPERKSEIFCRMDCVLSEWTEWS 470
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 2972 DSCNNCTCVSG----KMACTSER-----CFVA-----CGWSPWLWS 3004
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 471 SCQSQSNKNSDKQTRSRTILA-LAGEGKPCPPPSQALQEHRLNDHSC-----M 520
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 3005 LCSCSC---NVGIRRFPRAGTAPPAFAFGAEC--QQTWEAEFCSLRCPGPGVPGMCP 3058
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 521 QLHWETSPWGP--CSEDTLTALNAT-----345
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 3059 DKQWLDCAQFPASCABELSAPRGNTQCHPCHCPSGMILLVSRGHPGLGASVQPPVAL 3118
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 546 -----GW-----NGEATCGVIGITRRVFC-----VKSHV 569
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 3119 PGAIGTGSVEGAGWGPWPWSCHSRSCSGGLRSRTRACDPQPQGLGDYCEGPRAGCEV 3178
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 570 GOYMTKECP--DSTRPETYR--PCFLPKKDCIVTAFSEWTPCPRVCOAGNATVQKSYR 625
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 3179 CQALP--CPVTNCTALIEGAEYSFGPPPCRSDDLVHCVWR--CQPGCYCPGQVLSNGA 3235
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 626 IITQEAANGQCECPDLYBER-----ECEDVSLCPYRYRWKQP 662
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 3236 ICVQP---GHCSCLDLLTGORHHHPGARLARPDGNCNCTCLEGRINCTDLP-CPDC-GGGQ 3290
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 663 KWSPC-----ILVPSVWQ-GITGSEACG-----KGLQTRAVSCTISDDNESA 704
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 3291 SLHPCGQPCPERSCODLSPGSGVCPGSGVGCOPTCGCPGLQLSQDGLCVPPAHK----- 3342
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 705 EMMECLKQTINGM-PLLVQECTV----PCRE--DC---TFTAWSKTTPCSTNCG---EATK 750
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Qy	546	----	GW	-----	NGEATCGVGIQTRRVFC	-----	VKSHV	569					
Db	3119	PGAIGTGVGAGGWP	GWPSW	HCSSCGGLRS	TRACDP	PPQGLGLGYCEGP	RAQCEV	3178					
Qy	570	GOVMTKRCP	--DSTR	ETVYR	--PCF	LCKDCIVTAFSEW	TPCRVQAGNATV	QKSRYR	625				
Db	3179	COALP	--CPVNT	CTALEGAEY	SPCGP	PCRSDDL	VHCVMR	--QPGCYCP	PGQVLSNGA	3235			
Qy	626	IIIQEAANGQEC	PD	TLYBEER	-----	ECEDVSLCPYR	WKQP	662					
Db	3236	ICVQP	--GHCS	CLDLLTG	ORHHPGAR	LARPDCN	HCTCLEGR	LNCTDLP	--CPDC	GGGQ	3290		
Qy	663	KWSPC	-----	ILVPS	VMO	--GITGS	EAAG	-----	KGLQ	TRAVSCTISD	NRSA	704	
Db	3291	SLHFCGQPC	PCRS	QODL	SPGV	CVCGP	SGVGO	COPT	CGCP	LQGLSQDGL	CVPPAH	3342	
Qy	705	EMMECLUKQ	TNGM	--PLL	VQECTV	-----	PCRE	--DC	-----	TFTAWSK	FTPCSTNC	---EATK	750

b 3343 ----RCQYQAGAPSFVSTCVAGILQCOEVPDPPGVMWSWGWEDCSCVSCGGEQLR 3399  
y 751 SRR-----ROLTKSRKKEKQSDLYPLVET-ELCPD-EFISQPVGNWSD 795  
b 3400 SRRCARPPCGPARQRTCTQVCREAGCPAGRLYREQPGEGCFSCAHVTOVQVGFSE 3459  
y 796 ----CIIPEGRREHRLVQ-----ADSKCGEGLR 823  
b 3460 GCEBCHCEGTFQ-HRLACVQECPCVLTAMLLQELGATIGDQGPLGFGDELDSGQTLR 3518  
y 824 FRAVACSKNGR---PVPDPSFCSSGY-----IQEKCVIPCEP-- 858  
b 3519 TSCGNCSAHLKSLDDCFADGFGPWPSPGSCSGGLGTRTSRQCVLTMTPLS 3578  
y 859 ----DKLSDWSWGSSCSSGIGVIRSKWLKEKYNGRPPCKLIDLKNQ 905  
b 3579 ELFVCPGFCGAGNCSWTSWAPWEPSCRSYGVGOORRLAYR-PPGPGHWCNPILTAQ 3637  
y 906 VH-----EAVPCYSECNQYSWVVEHWSCKINNEL-RSLRCGGTQSR 947  
b 3638 ERRFCNLACPAAGCPAGNEVVTANRCPRCSDLOEGIVQDDQVQCGKCPKGSLEQ 3697  
y 948 KIRCV-----NTADGEG-----GAVDSNLQDEIPPETQSCSLM-C--PNECVMSWGL 994  
b 3698 DGGCVPIGHCDCTDAQHSHWAPGSHQDACNCSQAGQLSCTAQCPPPPHTCAWHSWA 3757  
y 995 WSKCPQSCDPHTMQRETRHLLPNSLR--TCAEDSOVQPCLL-----NE 1037  
b 3758 WSPCHSCGPRCQQRFR--CGFGLASRGSC-----PCLMAKADPTCNSTFLHLDTQ 3808  
y 1038 NCF-----QFOYNL-TWSTQLSNAPCGG--VTRLLSCVCSG--KPVSMDOCEQ 1086  
b 3809 GCYSGPCPDSCQSLGWPSPCQV---PCSGFRLRWREAEALCGGCRFPWAQDR--- 3861  
y 1087 HNLXPQMSIP--CLVRCV---NCLSG-WTAWTECSTQCHGHWSTRIIMPTQ 1139  
b 3862 ----KLORRALPSTCNVNSLVCPHOECPVLGFWPSWSSCSAPCG-GGTMERHRTC---EG 3913  
y 1140 GEG-RPCPELT-QBKTCPVTPCYSMVLGN-WSACKLEGGCGEGVQIRLSLM---VHS 1193  
b 3914 GGVAPCAQADTEQRECNLQPCPCPGQVLSAC-----ATSCPLCWLHQP 3961  
y 1194 GSISHAAGVEDALCGEPFQDSILKQISVPCPGDCHLTWSEWSTCELTCIDGRSFET 1253  
b 3962 G-----AICVQSPCPG-----CG--CPGGQSHLPWGLTLLE----- 3992  
y 1254 VGRQSRRTFIIQSFENODSCFQQVLETRPCTGKCYHYTWKASLNNNERTVNCQSRDG 1313  
b 3993 -----EQAQLPPGTVLTNRCTRCVCHGGAFCSL-----VDCQGE-- 4028  
y 1314 VNVTCGSPQAPAAIROCIAC-----RKPFYCTQGGVCGC 1351  
b 4029 IVPGETWQVAPGELGLCEQTCLMNAKTQSNCSARASGC 4071

RESULT 12

US-10-085-198-2  
Sequence 2, Application US/10085198  
Publication No. US2004000907A1  
GENERAL INFORMATION:  
APPLICANT: Alsobrook et al.  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-279  
CURRENT APPLICATION NUMBER: US/10/085,198  
CURRENT FILING DATE: 2002-02-25  
PRIOR APPLICATION NUMBER: 60/271,646  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 60/276,401  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/311,981  
PRIOR FILING DATE: 2001-08-13  
PRIOR APPLICATION NUMBER: 60/312,858

;; PRIOR FILING DATE: 2001-08-16  
;; PRIOR APPLICATION NUMBER: 60/271,840  
;; PRIOR FILING DATE: 2001-02-27  
;; PRIOR APPLICATION NUMBER: 60/277,324  
;; PRIOR FILING DATE: 2001-03-20  
;; PRIOR APPLICATION NUMBER: 60/286,096  
;; PRIOR FILING DATE: 2001-04-21  
;; PRIOR APPLICATION NUMBER: 60/299,695  
;; PRIOR FILING DATE: 2001-06-20  
;; PRIOR APPLICATION NUMBER: 60/315,614  
;; PRIOR FILING DATE: 2001-08-29  
;; PRIOR APPLICATION NUMBER: 60/272,405  
;; PRIOR FILING DATE: 2001-02-28  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 653  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 2  
;; LENGTH: 4219  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; US-10-085-198-2  
  
Query Match 8.0%; Score 655.5; DB 15; Length 4219;  
Best Local Similarity 21.9%; Pred.No.1.3e-41;  
Matches 390; Conservative 149; Mismatches 567; Indels 677; Gaps 101;  
  
QY 28 EQACLIPCRDCVSEFLPWSNCSKCGKGLQHRTEAVIAPPLFGGLQCP-NLTESRACD 86  
Db 2603 EDGCV-----DCVLAPWSVWSSCSRSGIGLTFQRELLRPPLPGG-SCPRDRFRSQC 2656  
QY 87 APISCPGEEYTFSLKVPWSKCRPLHLKEINPSRTVLDNFNSDSNERVTFKHOSYKAH 146  
Db 2657 VQ-ACFV-----AGAWAWERAWGPCSVSCG- 2681  
QY 147 HHSKSWAIGVQTRVQSTRSDGQ--AMLSLQDSPLTVQSCIMPKDCETSOWS-- 202  
Db 2682 -----CHQSRQSCVDPPPKNGGAPCPGASQERAPCGLOPCSGGTGKVLGWGHG 2731  
QY 203 -----SW-----SPCSKTCRS-----GSLLPGRFSRSRNVK 228  
Db 2732 GSTVGTGRGLPAPRLTWCPSTRLAGVCCEKSVPEGAGAMPILLPG--SQGQKL 2789  
QY 229 HMAIGG-----KCCPELLEKE----- 245  
Db 2790 QWALCGSLPSLLCPLGLSALFHLPGCRCPGGLLLHTRCLPLSECPCLVGBELKWPVS 2849  
QY 246 -----ACTIVE-CELLQO---CP-RYSWRT-SEWKEQVSLLEQDDPHWHVTGPVCGG 292  
Db 2850 FLGNCSCQVCEKCELLQCPGGCPLPCGWSAWSWAPCDRS-----CGS 2893  
QY 293 GIOTREYVCAQSVPAALALRAK-EVSRP-----VEKALCVGPAPLPSCQCNIP 339  
Db 2894 GVREAR--FRSPNPAPAAWGAPCEGDRQELQCHTVCGTGTAGSLGAGVPPSSQFCTLR 2951  
QY 340 C-----STDGIV-----SSWSANGLCIHENCHPQKKGKTRQHRVLMESTGPAGHC 387  
Db 2952 THGMGPTDHTWGIEVFGWTPMTWSSC--SQSLAPGGGFGWRSR--LCPSPDGSS--C 3007  
QY 388 P-HLVESVPCEDPM-CYRWLASSEGICFPD-----HGK----- 417  
Db 3008 PGDATCEPSPPIECTGFCAGCTCPGLFLHNASCLPRSCPCQLHGLQVYASGMARL 3067  
QY 418 -----CGLGHRILKAVQNDREGDVSGSLCFVPPPPPERKSCETPCRMDCVLSWTEWS 470  
Db 3068 DSCNNCTCVSG---KMACTSER-----CFVA-----CGWSPWTLWS 3100  
QY 471 SCQSOSCNKNSDGKOTRSRTILA-LAGEGKCPKPPSQALQEHRLCNDHSC-----M 520  
Db 3101 LCSGSC-----NVGIRRRFRAGTAPPAAFGAGSC--QGPTNEAEFCSLRCPGPFQMCPR 3154  
QY 521 QLHWETSPWGP--CSEDTLVTALNATI----- 545  
Db 3155 DKQWLCAQGPASCAELSAPRGNTQTHPGCHCPGSMILLVSPRHPGPGLASVQPPVAL 3214

546 -----GW-----NGEATCGVGIOTREVF-----VKSHV 569  
3215 PGAIGTGVPGAGGWPWGSWSRSTRACDQPPQGLGDBYCEBPQAQGEV 3274  
570 GOVWTKRCP--DSTRPTEVR--PCFLPKCKDCIVTAFSEWTPCPMCOAGNATVQSYR 625  
3275 CQALP--CPVTNCTAIEABYSFCGPPCRSCDDLHVCMR--CQPCYCPGPGVLSNGA 3331  
626 I11QEAANGQECPTLYBER-----ECEDVSLCPVVRWKPQ 662  
3332 ICVP-----GHCSCLDLTGRRHHPGARLPDGNCHCTCLEGRINCTDLP-CPDC-GGQ 3386  
563 KWSFC-----ILVPSVMQ-GITGSEAG-----KGLQTRAVSCISDNRSA 704  
3387 SLHPGQPCPRSCODLSPGVCVCGVGOCTCGPLGOLSDGLCVPPAHC----- 3438  
705 ENMECLKQTNM-PLVQECTV-----PCRE--DC-----TFTAWSKTPGCTNC--EATK 750  
3439 --RCQVQPCAMAPSFVPSTCVAGILQCOEVPDPPGVWSSWGPWEDCSVSGGGEQLR 3495  
751 SRR-----RQLTGKRKKEKQDSLDLYLVET-ELCPD-BFISQPVGNWSD 795  
3496 SRRCARPPCGPARQRTCTOVCREAGCPAGRLYRECQPGEGCPFCAHVTQVQCFSE 3555  
796 -----CILPGRREPHRLVQ-----ADSKGEGELR 823  
3556 GCEGCHPEGTQ-HLACVQPCVLTAWLLOELGATIGBPQPLGPDLDLSTQLR 3614  
824 FRVAGSDKNGR--PVDPSFCSSGY-----IOEKCVPICPP-- 858  
3615 TSCGNSCAHGKLSLDDCFEADGGFGPWPWGPCSRSGGLGTRSRQCVLWPTLS 3674  
859 -----DCKLSDWSGSSCGIGVIRSKWLKEKPYNGRPRCPKLDLKNQ 905  
3675 ELPVCPGPGGAGNCNWTSWAPWPCSRGCVGQORRLRAYR--PPGPGHWPNTLTAQ 3733  
906 VH-----EAVPCYSECNQYSVVVEHSSKINNEL-RSLRCGGGTQSR 947  
3734 ERRFCNLRAPEAGCPAGMEVTCANRCPRRCDLQEGIVCDDQVCQKGCPCPKSLEQ 3793  
948 KIRCV-----NTADGG-----CAVDSNLNODEIPETQSCSLM-C--PNECVMSWGL 994  
3794 DGGCVPIGHCDCTDAQHSAWPGSQHODACNCSOAGQLSCTAOPCPPTTHCAWSHA 3853  
995 WSKPOSQDPHTMQRTRHLLRSLNSR--TCAEDSQVQPCLL-----NE 1037  
3854 WSPCSHSGPRGQSRFR--CGFLASRSGC-----PCLMAKADPTCSNFTLHLDQ 3904  
1038 NCF-----QFQYNL-TEWSTCOLSENAFCGQ--VTRLLSCVCSG--KPVSDQCEQ 1086  
3905 GCYSGPCPDSQWSLWGPWSPQV-----PCSGGFLRWREALCGGCRPEWAQDR--- 3957  
1087 HNLEKFORMISIP-CLVCEVW-----NCQLSG-WTAMTECSQTCGHGRMSTRFTIMPTQ 1139  
3958 -----KLQRRALPCTCVNESLVCBQCPVLGWSANSSCSAPCG-GTWERHTC-----EG 4009  
1140 GEG-RPCPTLT--QKTCVTPYCSWLGN-WBACKLEGGDCGQVQIRLSLM---VHS 1193  
4010 GPGVAPCAQDTQRCQECNLQPCFPCPPGOVLSAC-----ATSCPLCWHLQP 4057  
1194 GSI SHAAGRVEDALCEMPQDSILKQLCSVPCPGDCHLTWSENSTCETLCIDGRSPET 1253  
4058 G-----AICVQEPQCG-----CG--CPGQHSPLWGLITLLE----- 4088  
1254 VGRQSRRTFIIOSFNQDSQCPQVLETRPCTGGKYHYTWKASLWNNHRTVWCQRSDG 1313  
4089 -----EQAQLPPTVLTNCTRCVCHGGAFCSL-----VDQGE-- 4124  
1314 VNYTGGCSQARPAARQCIAPAC-----RKPSYCTQGGVCCG 1351  
4125 IVPFGTWOQVAPGELGLCEQTCLMNATKTONCSARASGC 4167

RESULT 13  
US-09-972-467-2  
; Sequence 2, Application US/09972467  
; Patent No. US20020090373A1  
; GENERAL INFORMATION:  
; APPLICANT: PFIZER INC.  
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND  
; FILE OF INVENTION: US5 THEREOF  
; FILE REFERENCE: PCI0850A  
; CURRENT APPLICATION NUMBER: US/09/972,467  
; CURRENT FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1629  
; TYPE: PRT  
; ORGANISM: Human  
US-09-972-467-2  
  
Query Match 7.3%; Score 597.5; DB 9; Length 1629;  
Best Local Similarity 21.3%; Pred. No. 1.3e-37;  
Matches 225; Conservative 102; Mismatches 328; Indels 401; Gaps 50;  
  
QY 201 WSS---WSPCKTCRSGSLLPGRFSRNRNVKHAIGGKCEPILLEKEACIVGELL--- 254  
DB 881 WNSGHPQACSKPCQ-----GERKR-----KLVCTRSDQLTVS 914  
QY 255 -QCPRYSNRTSEWKEQVSLLEQQDPHWHVT-----GPVCGGIGTREVYCAQSVFAA 308  
DB 915 DQCRDLFPQGHITPCTGDC-----DLRWHASRSECSAQCGLYRTLDIYCAK----- 964  
QY 309 AALRAKEVSRPVEKALCVGAPLPSQLCNIPCTDCTIVSSW--SAGLGLCIHENCHEPQK 366  
DB 965 -YSLDGTKEVDGFC--SSHKPSN--REKSGECNCGWRYSANTEC----- 1009  
QY 367 KGPRTRORHVLMESTGPAGHCPHLVESVPCEDPMCVRWLASSEGICFPDGHGKGLHRLK 426  
DB 1010 -----SKSCDGGTQRRR 1021  
QY 427 AVCONDRGEDVSGSLCPVPPPPPKSC--EIPCRMDCVLSEWT--EWSSCSQSCSNKNSDG 483  
DB 1022 AICVNTNRDVLDDSKTHQEKVTIQRCEPFC-----PQWKGDMSECLVTC-----G 1069  
QY 484 KQTRSRITLALAGE--GGKPCFPFQALQEHRLCNDHSCMLHWETSPMGPCSEDTLVTA 540  
DB 1070 KGHKRVQWQCFGEDRLNDRMCDPETKPTSMQTCQPECAS--WQAGPMGQCS----- 1120  
QY 541 LNATIGNWGEATCGVGIOTRVEFC-VKSHVGOVMTKRCPDSTRTPEVTRPCFLPKKDC-- 597  
DB 1121 -----VTCGQYQLRAVKCIIGTYMSVVDDNDNNAATRTDTQCELP---SCHP 1167  
QY 598 -----IVTAFSEWTPCPMCOAGNATVQSRVRII1QEAANGQECPTD 641  
DB 1168 PPAAPETRTSTYSAPRTQWRFSGSWTPCSATCGKG-----TRVRYVSCRDENG-----S 1215  
QY 642 LYBEREC-----EDVSLCPVYWKPKQWSPCLVLPESVWQGITGSEACGKLOTR 692  
DB 1216 VADESACATLPRPVAKEECSVTFCGQWALDWSSC-----SVTCQGRATR 1261  
QY 693 AVSCISDNRSAEMMECLKQTNMGVPLLVQECTVPCREDCTFTAWSKFTPCSTNCEATKSR 752  
DB 1262 QVMCVNYSVHVIDRSEC--DQDYIPETDQDCSM-----SPC----- 1295  
QY 753 RRLQTKRKKKCKQSDLL--YPLVETELCPDDEFTISQPY-----GNWSDCILPEG 801  
DB 1296 -----PORTPSGLAQHPFQNEQYRPSASPSRTHVLGNGQWRTGPMGAC----- 1340  
QY 802 RREPHRGLRVQADQKCEGEGELRFRVACSDKNG-----RPVDPSPCSSSGVIOEK 851  
DB 1341 -----SSTCAGSQRRVVVQDENGYTANDVERIKPDEQACESG----- 1381  
QY 852 CVTIPCPFDCKLSDWS--SWGSCSSSCGIGVIRSKWLKEKPYNGRPRCPKLDLKNQVHEA 909

1382 ---PCP-----QWYGNWGBCTKCGGIRTR---LVVQORSNGERFP--DLSCILDK 1427  
910 VPCVSECNQY-----SWVHEWSSCKINNELSLRCLCGGTQSRKIRCVNTADGEGAVD 963  
1428 PPDRECCNTHACPHDAASTGSPWSC-----SVSCGRGHQRNVYCMAK---DGSHELE 1477  
964 SNLCNQDBIPPETQSCSLMCPNECVMSWGLWSKCPQSCDPHTWQRRTHILLRPSLSRT 1023  
1478 SDYCKHLAKPHGRKCR-----GGRCP-----1499  
1024 CAEDSQVOPCLLNENCFQYNLTWSTCOLSENAPCGGVTRILLSCVSDGKPVSMQDQ 1083  
1500 -----KMKAGAWSOCVS-----CGRGVQORHVGCQIOTHKIARETE 1536  
1084 CEHNLKRPQMSIPCLVECV-VNQLSGWTA--WTECSQTCGHGGRMSRTRFIIM-----1136  
1537 CNPYTRPESR-----DCQGRCPFLYTRAEWQECTKTCGEG---SRYRKVVVVDN 1586  
1137 PTQEGRCPTTEL--TQKTCPTVPC-YSWULGNWS 1169  
1587 KNEVHGARCVDKRPVDRESCSLQPCETVWITGWS 1622

## RESULT 14

S-09-796-753-20  
Sequence 20, Application US/09796753  
Publication No. US20030027998A1  
GENERAL INFORMATION:  
APPLICANT: McCarthy, Sean A.  
TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF  
FILE REFERENCE: 7853-227-999  
CURRENT APPLICATION NUMBER: US/09/796,753  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 09/183,175  
PRIOR FILING DATE: 1998-10-30  
PRIOR APPLICATION NUMBER: 09/223,094  
PRIOR FILING DATE: 1998-12-30  
PRIOR APPLICATION NUMBER: 09/223,546  
PRIOR FILING DATE: 1998-12-30  
PRIOR APPLICATION NUMBER: 09/224,246  
PRIOR FILING DATE: 1998-12-30  
PRIOR APPLICATION NUMBER: 09/259,388  
PRIOR FILING DATE: 1999-02-26  
PRIOR APPLICATION NUMBER: 60/122,458  
PRIOR FILING DATE: 1999-03-01  
PRIOR APPLICATION NUMBER: 09/312,359  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 09/336,536  
PRIOR FILING DATE: 1999-06-18  
PRIOR APPLICATION NUMBER: 09/342,687  
PRIOR FILING DATE: 1999-06-29  
PRIOR APPLICATION NUMBER: 09/345,464  
PRIOR FILING DATE: 1999-06-30  
PRIOR APPLICATION NUMBER: 09/365,164  
PRIOR FILING DATE: 1999-07-30  
PRIOR APPLICATION NUMBER: 09/399,723  
PRIOR FILING DATE: 1999-09-20  
PRIOR APPLICATION NUMBER: 09/409,634  
PRIOR FILING DATE: 1999-09-30  
PRIOR APPLICATION NUMBER: 09/471,179  
PRIOR FILING DATE: 1999-12-23  
PRIOR APPLICATION NUMBER: 09/474,071  
PRIOR FILING DATE: 1999-12-29  
PRIOR APPLICATION NUMBER: 09/474,072  
PRIOR FILING DATE: 1999-12-29  
PRIOR APPLICATION NUMBER: 09/514,010  
PRIOR FILING DATE: 2000-02-25  
PRIOR APPLICATION NUMBER: 09/516,745  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 09/572,002  
PRIOR FILING DATE: 2000-05-14  
PRIOR APPLICATION NUMBER: 09/597,993

PRIOR FILING DATE: 2000-06-19  
PRIOR APPLICATION NUMBER: 09/599,596  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/630,334  
PRIOR FILING DATE: 2000-07-31  
PRIOR APPLICATION NUMBER: 09/606,565  
PRIOR FILING DATE: 2000-06-29  
PRIOR APPLICATION NUMBER: 09/606,317  
PRIOR FILING DATE: 2000-06-29  
PRIOR APPLICATION NUMBER: 09/665,666  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: 09/677,751  
PRIOR FILING DATE: 2000-09-30  
NUMBER OF SEQ ID NOS: 162  
SEQ ID NO 20  
TYPE: PRT  
LENGTH: 874  
ORGANISM: Homo sapiens  
US-09-796-753-20  
Query Match 6.2%; Score 507.5; DB 10; Length 874;  
Best Local Similarity 20.1%; Pred. No. 6.8e-31;  
Matches 231; Conservative 126; Mismatches 332; Indels 463; Gaps 53;  
QY 201 WSSWSPCSKTCRSG---SLLPGFRSRS---RVYKMAIGGKCEPCL---LEKEACIVBG 251  
Db 39 WGPWSECSRTCGGASYSLRCLSSKSCGRNIRYTC-SNVDCPEAGDFRAQCSAHN 97  
QY 252 ELQOCQPRYSWRTSEWKECQVSLLEQQDPHMHVTGFCVGGGIQTRVYCAQSVFAAAL 311  
Db 98 DVKHGQFYEW-----LPVSNDF-----DNPCLAK 123  
QY 312 RAKEVSRPVEKALCVCPALPQLCNIPGSTDIYSSWSAWGLCIHENCHEPQKKGFT 371  
Db 124 QAKGTTLVVELA-----PKVLDGTRCYTESL-----DMCISGLC-----157  
QY 372 RQRHVLMESTGPAGHCPHLVESVPCBDPMCYWLASEGICFPDHGKCGHGRILKAVCON 431  
Db 158 -----QIVG---CDHQLGSTVKED-----NCGVCNGDGSCTCLVRGQYKQLSA 198  
QY 432 DRGEDV-----SGSLCPVPPPPPERKSCIEICRMDCVLSEWTEWSSCSQCSNKNDSK 484  
Db 199 TKSDDTVAIPYGRSHIRLVKGPDLHYLETK-TLQGTGENSELSSTGTFLVDNSVDQ 257  
QY 485 QTRSRTILALAGEGKPCPPSOALQEHRLCNDHSCML-----HWETSFWGCSDEDT 536  
Db 258 KFPDKELRMAG---PLTADPIVKIRNSGSADSTVQFYQPIHRWEETDFPCS---310  
QY 537 LVTALNATIGWGEATCGVGIGTRRVFCVKSHVQVMTRK-C---PDSTRPTVTRPCFLP 592  
Db 311 -----ATCGGGYQLTSAECYDLRSNRVADQYCHYYPENIKP---KPKLOE 353  
QY 593 CKKDCIVTAFSEWTPCPRMCOAGNATVKOSRYRIIQEAANGGQECPTLYERECEDEYS 652  
Db 354 CNLD-----PCP-----AQRVGIIVPNSDGYKQIMPYLYH-----384  
QY 653 LCPVYRWKPKQWSPCILVPESVMVQGITSGSEACGKLQTRAVSCISDDNR-----SABMME 708  
Db 385 --PLRWEATPWTAC-----SSSCGGGIQSRVSCVEEDIQGHVTSVEEWK 428  
QY 709 CLAQTNMPLLYQECTVPCREDCTFTAMSKFTPCSTNCEATKRRRLTGRKKEKCOD 768  
Db 429 CM-YTPRMPIA-----QPCN-----442  
QY 769 SDLYPLVETELCPDEFISQPYGNWSDCILPEGRBPHRGLRVQADSKCEGGLRFRAYA 828  
Db 443 -----IPDCPKWLAQ---EWSPCIV-----TCGQLRVRVL 471  
QY 829 CSDKNGRPVDPFCSSSGYIOEKVIP-----CPFDCKLSDW-----865  
Db 472 CIDHRCGMHTGGCSPKTKPHKECIVPTPCYKPKELPVEAKL-PMFQKQALEGAAVS 530  
QY 866 -----SSWGSCTSSCGIGVIR-----SKWLKEKPYN---GGRPCPKLDLKNQ 905

```

Db 531 EEPSPFPEAMSACTVTCGVTQVIRVRCQVLLSFSQSVADLPIDECEGPKPA-----SOR 585
2Y 906 VHEAVPCYSCNQ-----YSWVVEHWSCKINNELSLRSCGGTOS 946
Db 586 ACVAGPCSGEIPFNPDETDLGEGQLDFDELVDWEYEGFTK-----SSECGGVQOE 638
2Y 947 RKIRCVNTADGEGAVDSNLGNODEIPPE--TQCSLWMCNPEVMSWGLWSKQSCDPDH 1005
Db 639 AVVSLCNKQTR--PAEENLCVTSRRPPQLKSCNL-----DPCPA-----677
2Y 1006 TMOQRTRHLLRPSLNSRTCAEDSQVQPCLLNENCFOFYNNLTWSTCOLSENAPCGQYR 1065
Db 678 -----RWEIGKWSPCSLT-----CVGGLQ 696
2Y 1066 TRLLSC-----VCSGDKPVSMDOCEQHNLEKRPQMSIPCLVECVVNCQ 1108
Db 697 TRDVFCSHLLSREMNETVILADELCRQPKPSTVOACNR-----FNCP 738
2Y 1109 LSGHTA-WTECSQTCGHG-----GRMSRTRFIIMPTQ--GEGRPCTELTQKTC 1156
Db 739 PAMYPAGQPCSRCTCGGVQKREVLCKQRMADGSFLELPETFCASQACQACKKDDCP 798
2Y 1157 VTPCYSVLGNWSACKLEGDCGEGVQIRSLSC--MVHSGSIHAAAGRVEDALCGEMPFQ 1214
Db 799 ----SEWLLSDWTECST---SCGEGTQTRSAICRKMKTG-ISTV---VNSTLCPLPFS 847
2Y 1215 DSILKQLCVP 1226
Db 848 SSIR-----PC 853

RESULT 15
US-10-044-807-2
; Sequence 2, Application US/10044807
; Publication No. US20020165187A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; TITLE OF INVENTION: Turner, C. Alexander Jr.
; FILE REFERENCE: LEX-0298-USA
; CURRENT APPLICATION NUMBER: US/10/044,807
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 60/261,684
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1762
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-044-807-2

Query Match 6.0%; Score 496; DB 13; Length 1762;
Best Local Similarity 19.8%; Pred. No. 1.4e-29;
Matches 228; Conservative 127; Mismatches 324; Indels 474; Gaps 53;

2Y 201 WSWSPCKTCRSG---SLPFGFSRS---RNVKHMAIGGKSCPEL---LEKXACTIVEG 251
Db 39 WGPWSECSRTCGGASYSRLRCLSSKSCGEGNIRYRTC-SNVDCPPRAGDFRAQQCSAHN 97
2Y 252 ELLAQCPYRSWRTSEWKECVSLLEQQDPHHVYTGVPVCGGIGTREVYCAQSVPAAL 311
Db 98 DVKHGQFYEW-----LPVNDP-----DNPCLSK 123
2Y 312 RAKEVSRPVEKALCVGAPAPLSQJCNIPCTDCTVSSWSAWGLCIHNCHEPOCKGFR 371
Db 124 QAKGTTLVLA-----PKVLDGTRCYTESL-----DMCISGLC-----157
2Y 372 RQRHVLWESTPGAGHCHPLVESVPCEDPMCVRLASRGICFPDGHGKGLGHLKAVCON 431
Db 158 -----QIVG-----CDHQLGSTVKED-----NCGVNGDSTCLRVGQYKLSA 198

```

```

QY 432 DRGEDV-----SGSLCPVPFPPPERKSCFIPCMDCVLSEWTESSCSQSCSNKSDGK 484
Db 199 TKSDDTVVAIPYGSRHRLVLKGPDLHLYLETK-TLOGTGKENSLSSTGTFLVDNSSVDFQ 257
QY 485 QTRSDTILALAGEGKCPSPSOALOEHLNDSHCOL-----HWETSPWGPCSEDT 536
Db 258 KFPDKELRMAG-----PLTADFIKIRNSGASSTVQFIYQPIIHRWRETDFPFS 310
QY 537 LVTALNATIGNWGATCGVGIQTRRVFCVKSHVQVMTKR-C---PDSTRPVTVPCLP 592
Db 311 -----ATCGGGYQLTSAECYDLRSNRVADQVCHYYPENIKP---KPKLQE 353
QY 593 CKDCIVTASEWTPCRMCAQGNATVQKSYRIIIQEAANG-GOECPDLYEBERECEDV 651
Db 354 CNLD-----PCP-----ASDGYKQIMPYDLYH-----375
QY 652 SLCPVYRWKPKOKSPCILVPESVMQGITSGSEACGKGLQTRAVSCISDDNR-----SAEMM 707
Db 376 ---PLPRWEATPWATAC-----SSCGGIGQSRVSCVVEEDIQGHVTSVEEW 418
QY 708 ECLQQTNGMPLLVQECTVPCREDCTFTAWSKFTPCSNCEATKSRRLQITKSRKKEKCQ 767
Db 419 KCM-YTPKMPITA-----QPCN-----433
QY 768 DSDLVPLVETELCPDCEFIQPYGNWSDCILPEGRREPHRLRVQADSKECCEGLRPAV 827
Db 434 -----IFDQPKWLAQ---EMSPCTV-----TCGQLRYRVV 461
QY 828 ACSKNGRPVDPSPCSSSGYIOEKCVP-----CPFDCKLSDW-----865
Db 462 LCIDHRGMHTGCGSPKTKPHIKEECIVPTCYKPEKLPEAKL-PWFKQAQAELEGA 520
QY 866 -----SSWGSCTSSCGIGVIRIR-----SKWLKEKPNY---GGRPCPKLDKN 904
Db 521 SEEPSFPEAMSACTVTCGVTQVIRVRCQVLLSFSQSVADLPIDECEGPKPA-----SQ 575
QY 905 QVHEAVPCYSCNQ-----YSWVVEHWSCKINNELSLRSCGGTQ 945
Db 576 RACYAGPCSGEIPFNPDETDLGEGQLDFDELVDWEYEGFTK-----SSECGGVQ 628
QY 946 SRKIRCVNTADGEGAVDSNLGNODEIPPE--TQCSLWMCNPEVMSWGLWSKQSCDP 1004
Db 629 EAVVSLCNKQTR--PAEENLCVTSRRPPQLKSCNL-----DPCPA-----668
QY 1005 HTMQRTRHLLRPSLNSRTCAEDSQVQPCLLNENCFOFYNNLTWSTCOLSENAPCGV 1064
Db 669 -----RWEIGKWSPCSLT-----CGVGL 686
QY 1065 RTLELLSC-----VCSGDKPVSMDOCEQHNLEKRPQMSIPCLVECVVNC 1107
Db 687 QTRDVFCSHLLSREMNETVILADELCRQPKPSTVOACNR-----FNC 728
QY 1108 QLSGHTA-WTECSQTCGHG-----GRMSRTRFIIMPTQ--GEGRPCTELTQKTC 1155
Db 729 PAMYPAGQPCSRCTCGGVQKREVLCKQRMADGSFLELPETFCASQACQACKKDDC 788
QY 1156 PVPYCYSVLGNWSACKLEGDCGEGVQIRSLSC--MVHSGSIHAAAGRVEDALCGEMPF 1213
Db 789 P-----SEWLLSDWTECST---SCGEGTQTRSAICRKMKTG-----LSTVNSTLCPLP 837
QY 1214 QDSILKQLCVP 1226
Db 838 SSIR-----PC 844

```

Search completed: February 24, 2004, 01:30:05  
Job time : 619.973 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 23, 2004, 10:50:55 ; Search time 1592.77 Seconds  
(without alignments)  
11538.043 Million cell updates/sec

Title: US-10-022-710-5\_COPY\_1\_424

Perfect score: 424  
Sequence: 1 aggttccaaagacaaact.....gcatggactgcagacacgga 424

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pin.\*
- 35: em.htg.rod.\*
- 36: em.htg.man.\*
- 37: em.htg.vrt.\*
- 38: em.sv.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	294.4	69.4	6112	10	BC063250	BC063250 Mus muscu
C 2	290.4	68.5	119205	9	AC011231	AC011231 Homo sapi
C 3	290.4	68.5	167462	2	AC027600	AC027600 Homo sapi
C 4	209	49.3	5724	9	AB051466	AB051466 Homo sapi
C 5	200	47.2	224789	2	AC111611	AC111611 Rattus no
C 6	195.2	46.0	259236	2	AC123741	AC123741 Mus muscu
C 7	141	33.3	164366	2	AC079808	AC079808 Homo sapi
C 8	141	33.3	180602	9	AC010873	AC010873 Homo sapi
C 9	110.6	26.1	234528	2	AC096222	AC096222 Rattus no
C 10	104.2	24.6	175131	2	AC124806	AC124806 Mus muscu
C 11	78.6	18.5	3053	6	AX747433	AX747433 Sequence
C 12	78.6	18.5	3053	9	AK092252	AK092252 Homo sapi
C 13	76	17.9	6378	6	AR338841	AR338841 Sequence
C 14	75.4	17.8	6373	6	AX079870	AX079870 Sequence
C 15	75.4	17.8	112456	9	AC004614	AC004614 Homo sapi
C 16	75.4	17.8	170419	9	AC146265	AC146265 Pan trogl
C 17	75.4	17.8	184549	2	AC068575	AC068575 Homo sapi
C 18	73.8	17.4	187347	2	AC135033	AC135033 Rattus no
C 19	73.8	17.4	234459	2	AC127064	AC127064 Rattus no
C 20	70.6	16.7	172164	2	AC115892	AC115892 Mus muscu
C 21	36.6	8.6	275073	2	AC115176	AC115176 Rattus no
C 22	36.4	8.6	238386	2	AC109763	AC109763 Rattus no
C 23	36.4	8.6	345974	2	AC110656	AC110656 Rattus no
C 24	36.2	8.5	4320	10	AK129133	AK129133 Mus muscu
C 25	36.2	8.5	8456	10	AB072381	AB072381 Mus muscu
C 26	36.2	8.5	215581	2	AC106161	AC106161 Rattus no
C 27	36.2	8.5	230955	2	AC095647	AC095647 Rattus no
C 28	36.2	8.5	235977	2	AC132507	AC132507 Rattus no
C 29	35.8	8.4	1629	6	AX167247	AX167247 Sequence
C 30	35.8	8.4	3460	10	BC031804	BC031804 Mus muscu
C 31	35.8	8.4	166743	10	AC127226	AC127226 Mus muscu
C 32	35.8	8.4	182921	10	AC124761	AC124761 Mus muscu
C 33	35.8	8.4	183707	10	AC124384	AC124384 Mus muscu
C 34	35.6	8.4	163209	9	AC093419	AC093419 Homo sapi
C 35	35.4	8.3	132169	2	AC026814	AC026814 Mus muscu
C 36	35.4	8.3	175144	2	AC068502	AC068502 Mus muscu
C 37	35.4	8.3	209857	10	AL583883	AL583883 Mouse DNA
C 38	35.2	8.3	124831	2	AC121270	AC121270 Mus muscu
C 39	35.2	8.3	21452	2	AC121752	AC121752 Rattus no
C 40	35	8.3	11505	1	AE001936	AE001936 Deinococc
C 41	35	8.3	124256	2	AC124151	AC124151 Oryza sat
C 42	35	8.3	166300	9	AC018884	AC018884 Homo sapi
C 43	35	8.3	179692	2	BX539346	BX539346 Danio rer
C 44	35	8.3	185773	2	AC027528	AC027528 Homo sapi
C 45	35	8.3	206214	2	AC079859	AC079859 Homo sapi

ALIGNMENTS

RESULT 1  
BC063250  
LOCUS BC063250  
DEFINITION Mus musculus RIKEN cDNA D130067103 gene, mRNA (CDNA clone MGC:166597 IMAGE:6410178), complete cds.  
ACCESSION BC063250  
VERSION BC063250.1 GI:38649124  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 6112)  
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L., Schaefer, T.E., Brownstein, M.J., Udell, R.B., Tohyama, Y., Schaefer, T.E., Prange, C., Rana, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Bickerton, Y.S., Krzywicki, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.B., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)

12477932

2 (bases 1 to 6112)

Strausberg, R.

Direct Submission

Submitted (02-DSC-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgapbe@mail.nih.gov](mailto:cgapbe@mail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>

Series: IRAP Plate: 125 Row: 1 Column: 17

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 27369657.

Location/Qualifiers

1..6112

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="MGC:66597 IMAGE:6410178"

/tissue\_type="Brain, enriched mouse brain 12.5dp"

/clone\_lib="NIH\_BMAP\_F00"

/lab\_host="DH10B"

/notes="vector: pYX-ASC"

1..6112

/gene="D130067I03Rik"

/note="synonym: B930082A18"

/db\_xref="LocusID:210417"

/db\_xref="MGI:2443925"

162..4985

/gene="D130067I03Rik"

/product="D130067I03Rik protein"

/protein\_id="AAH63250.1"

/db\_xref="GI:38649125"

/db\_xref="LocusID:210417"

/db\_xref="MGI:2443925"

/translation="MFLRSLAVTHVSRMRKFLVLSLLLSQAHLRGRKDNQFLW KTGPMGACDGGPGGSAVWCFHLEGWTSFMSNCDESSQPPKRSFRVCDWHSDFQWEDVHRCFLVAPGQGEPRVAVCVTAQHLQHRTRVCLQKLNRTMVSNEICE HFAPQPTQACILPCPRDCVBSFSPWSTCPGEGCKGLQHRTRVAIAPPLYGLQCP

NTLSRACEAPVSPGLGKEEVSFLKVPWSKCRLLPHLKEVDLSGRNIQDFSSDSNEQ  
VTLTHQSVKAHHSQPDGVITGFOTVACTSRTSNGRALLSLCYRDSFPLTVQBCVME  
KDCSTSEHSPMSKCTCRSGTILSPGVRSRSRNKHIAIGCGQCPRELLEKTECIAEG  
EPLQCPRIYSEWKECVSLLELQHDPLWHETGPIGGIQTIREVYICQSLPALT  
ASRTKEVRPVELTICLGPAPASOLCNVPCSMDCIVSSWSGTGFCFPCFENCHDPOGKK  
GFRQRHVMZMETGPMCRPLAESVPCEDPMCRWLASEGICIAHKGKGLGKLRSDG  
KAVQNERGEVSGGLCPVPPPPPMACEIPCRVDCVSWVTWSSSQSCSNKNSDL  
QKTRSRSLALAGSGGLCTPSSQELVRLCNDHSTCTQLYWETWASGSCSNTLVATL  
NVTIGMGEATCGVIGTQTRKVCIKSHVGOVMTKRCPESTETVTRPCFLPKCKDCLV  
TAFSEWTPCPHPCOPGPTIKOSYVRIILQEAANGGOCEDPTLPEERCECISLCPSY  
RWKPKWSCHILPESIRQGRITSEACGKGLQTRAVSCISDDNQSRHMTISLQKMG  
LPELVQECTIFRDDCTIPMSKFTPSKNCETATQIRRRQTKGSKRCKQDASLYP  
MMEVPCDCTFMGHPHYNGSACILPEKGRDAQGGGLWQDGDKECEGVEFRALIAS  
NINRPGDPSFNSGTYQEAICPCFPDCKLSDWSWSGSCSSCGVIRISRWLKE  
KPYSGRCPKLDLKNQVHEAVPCYSECQDSWVVEHWSCKINNEIRSPRCGRGTQS  
HAVORTEHILRPSINRSTCGEDSQVRPCILNCEFOYNLTJMSCTQISENVSCGQ  
GYRRLLSCVASDGSVMDHCEORNELEKFORMSIPCLVECVNCLQSGTWTTECSQ  
TCQGGWRSTRFILMPTQEGROCPTELQKPCPVTPTCYWVLGWSACKISGGDC  
GEGVQVRSFCVHNGSISHTAVPVEALGVEFPQEGILKQLCSPVCPGCHITPWS  
EWSKELCTIDGRSFETTRQSRRTTIOSEFENDSCPPQVLETRTCTGKCYHYTW  
KASLWNNERTWQCRDGLNVTGCGSPQAPAAIROICIPACKPFYSYCTGGVCGCE  
KGYTEIMSSGFLDYCMKVPGESDKADVKNLGNKRPVNSKIHIIPKWSLQLOLDPD  
GRVXWVWYVSGSGSLMIFLVFTSYLVCKXPKPHQSTPRHQKPLTLAYDGLDM"

705..842

/gene="D130067I03Rik"

/note="TSPI; Region: Thrombospondin type 1 repeats"

/db\_xref="CDD:smart00209"

1179..1358

/gene="D130067I03Rik"

/note="TSPI; Region: Thrombospondin type 1 repeats"

/db\_xref="CDD:smart00209"

3915..4067

/gene="D130067I03Rik"

/note="TSPI; Region: Thrombospondin type 1 repeats"

/db\_xref="CDD:smart00209"

ORIGIN

Query Match 69.4%; Score 294.4; DB 10; Length 6112;

Best Local Similarity 80.9%; Pred. No. 4e-83;

Matches 343; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGAGCAACTAACAGTCACTTCTGGGTATGGAGGAGCATGAGGAAGTC 60

Db 162 ATGTTTCTGAGGAGTGACTGCTGCTCACTCACTGGGTTTCCAGGAGCATGAGGAAGTC 221

QY 61 TTTCTATTGCTTTCTCTTCTGCTCCATGCAGTCACTTGGAGGAGCAAAAGATAAT 120

Db 222 TTTCTAGTGTCTTTCTCTCTGCTCTCCCAAGCAGTCACTTGGAGGAGCAAAAGATAAT 281

QY 121 CAGTTTCATCTGGAACACAGGTCCTGGGAGAGGTGTACAGGAGACTGTGGTCCCGGAGGA 180

Db 282 CAGTTTCCTCTGGAACACCGTCCCTGGGAGAGGTGTGAGTGTCTGGGACGAGGGA 341

QY 181 GTCCAGAGTCCGGGAGTGTGTGTTTTCATGTTGACGGGTGGACAAGTCACTGTCTAAC 240

Db 342 GCCCAAGTCTGTGTATGTGTGTTTTCACATAGAGGGGTGGAGTCCCATGTCCGAAC 401

QY 241 TGTGTGTAGAGCAACAGAGCCCTCCAAAGAGAGAGTGTGTTTCCGAGTTTGTGACTGGCAC 300

Db 402 TGTGTATGAGAGCAGCCACCTCCAAAGAGAGAGTGTGTTTCCGCGTGTGTGACTGGCAC 461

QY 301 AGTGACCTCTTTTCAAGTGGAGGTTTCTGACTGGCACCACCTGTGTGCTTGTTCCTTACGCT 360

Db 462 AGTGACCTGTTCAGTGGAGGTTTCCGACTGGCATCGCTGCCCTCTGTGTTTCTTGGGCGC 521

QY 361 CGCGGTGAAGTCAGAGCTCGAGCTGCAGGTGTGTGAGGTGTGAGGTGTGAGTGTGAGTGTG 420

Db 522 CAAAGTGAAGCCAGGCTTCGGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 581

QY 421 CGGA 424

Db 582 CGGA 585





```

repeat_region /rpt_family="L1"
19109..19386
repeat_region /rpt_family="Alu"
19387..19705
repeat_region /rpt_family="L1"
19706..19870
repeat_region /rpt_family="(TA)n"
19890..20882
repeat_region /rpt_family="L1"
20883..21186
repeat_region /rpt_family="Alu"
21187..21558
repeat_region /rpt_family="L1"
21559..21583
repeat_region /rpt_family="(TTT)G)n"
21584..21872
repeat_region /rpt_family="L1"
21950..21984
repeat_region /rpt_family="AT-rich"
22011..22261
repeat_region /rpt_family="L1"
22474..22706
repeat_region /rpt_family="WAlR"
22814..23079
repeat_region /rpt_family="Alu"
23190..23461
repeat_region /rpt_family="L1"
23518..24159
repeat_region /rpt_family="L1"
24270..24652
repeat_region /rpt_family="WAlR"
24696..24959
repeat_region /rpt_family="L1"
26424..26590
repeat_region /rpt_family="MIR"
27884..28177
repeat_region /rpt_family="L1"
28632..28930
repeat_region /rpt_family="Alu"
28940..28990
repeat_region /rpt_family="(CATATA)n"
29009..29317
repeat_region /rpt_family="Alu"
29554..29981
repeat_region /rpt_family="L1"
29982..30099
repeat_region /rpt_family="Alu"
30100..30583
repeat_region /rpt_family="L1"
30826..32314
repeat_region /rpt_family="L1"
38581..38608
repeat_region /rpt_family="AT-rich"
38609..38984
repeat_region /rpt_family="L1"
39073..39148
repeat_region /rpt_family="L1"
39214..39340

Query Match 68.5%; Score 290.4; DB 9; Length 119205;
Best Local Similarity 86.3%; Pred. No. 9.9e-82;
Matches 321; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 53 GGAGCTCTTTCTATTGTTCTCTTTGTCCTCCATGAGCTCATTTGGAGGCAAAA 112
DB 52812 GGAAAGTGTGTAATGACTTCTACTTACCTGTGTAATGCTGTAATCTCTGAGTAA 52753
QY 113 AGGATAATCAGTTCATCTGGAACACAGTCCGTGGGAAGGTACAGGAGACTGTGTC 172
DB 52752 TTACATCTCTGTTTCTGCTGTAGTCTCGTGGGAAGGTACAGGAGACTGTGTC 52693
QY 173 CCGGAGGAGTCCAGAGTCCGGCAGTGGTGTGTTTTCATGTTGACGGTGGACAGTCACC 232

```

```

Db 52692 CCGGAGGAGTCCAGAGTCCGGCAGTGTGTTTTCATGTTGACGGTGGACAGTCACC 52633
QY 233 TGTCTAACTGTGGTGAGAGCAACAGGCTCCAAAGGAAAGAGTGTGTTCCGAGTTTGTG 292
Db 52632 TGTCTAACTGTGGTGAGAGCAACAGGCTCCAAAGGAAAGAGTGTGTTCCGAGTTTGTG 52573
QY 293 ACTGGCAGAGTGAACCTCTTTTCAGTGGGAGGTTTCTGACTGGCACCACCTGTGTGTTTC 352
Db 52572 ACTGGCAGAGTGAACCTCTTTTCAGTGGGAGGTTTCTGACTGGCACCACCTGTGTGTTTC 52513
QY 353 CTTAGCTTCGGGTTGAGTCAAGCTCGACTCGAGTGTGTCAGCGCTCAGCATGGAC 412
Db 52512 CTTAGCTTCGGGTTGAGTCAAGCTCGACTCGAGTGTGTCAGCGCTCAGCATGGAC 52453
QY 413 TGCAGCACCGGA 424
Db 52452 TGCAGCACCGGA 52441

RESULT 3
AC027600/c
LOCUS
DEFINITION 167462 bp DNA linear HTG 07-JUN-2000
SEQUENCE, 22 unordered pieces.
ACCESSION AC027600
VERSION AC027600.2 GI:8318467
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 167462)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 2, clone RP11-41418
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 167462)
AUTHORS Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marcuis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivaz,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodores,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
JOURNAL Direct Submission
COMMENT Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 7, 2000 this sequence version replaced gi:7342347.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9157
Center clone name: 414_I_8
----- Summary Statistics

```

Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 156843 bases at least Q40  
Consensus quality: 161829 bases at least Q30  
Consensus quality: 163971 bases at least Q20  
Insert size: 170000; agarose-fp  
Quality coverage: 4.2 in Q20 bases; agarose-fp  
Quality coverage: 4.3 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1293: contig of 1293 bp in length  
\* 1294 1393: gap of 100 bp  
\* 1394 3228: contig of 1835 bp in length  
\* 3229 3329: gap of 100 bp  
\* 3329 5140: contig of 1812 bp in length  
\* 5141 5240: gap of 100 bp  
\* 5241 8017: contig of 2777 bp in length  
\* 8018 8117: gap of 100 bp  
\* 8118 10466: contig of 2349 bp in length  
\* 10467 10566: gap of 100 bp  
\* 10567 13763: contig of 3197 bp in length  
\* 13764 13863: gap of 100 bp  
\* 13864 15896: contig of 2033 bp in length  
\* 15897 15997: gap of 100 bp  
\* 15998 19987: contig of 3991 bp in length  
\* 19988 20087: gap of 100 bp  
\* 20088 24063: contig of 3976 bp in length  
\* 24064 24163: gap of 100 bp  
\* 24164 26824: contig of 2661 bp in length  
\* 26825 26924: gap of 100 bp  
\* 26925 30928: contig of 4004 bp in length  
\* 30929 31028: gap of 100 bp  
\* 31029 34504: contig of 3476 bp in length  
\* 34505 34604: gap of 100 bp  
\* 34605 40409: contig of 5805 bp in length  
\* 40410 40509: gap of 100 bp  
\* 40510 46642: contig of 6133 bp in length  
\* 46643 46742: gap of 100 bp  
\* 46743 55317: contig of 8575 bp in length  
\* 55318 55417: gap of 100 bp  
\* 55418 64038: contig of 8621 bp in length  
\* 64039 64138: gap of 100 bp  
\* 64139 71781: contig of 7643 bp in length  
\* 71782 71881: gap of 100 bp  
\* 71882 81518: contig of 9536 bp in length  
\* 81519 81518: gap of 100 bp  
\* 102058 102157: contig of 20540 bp in length  
\* 102158 102157: gap of 100 bp  
\* 102158 121599: contig of 19442 bp in length  
\* 121600 121699: gap of 100 bp  
\* 121700 142729: contig of 21030 bp in length  
\* 142730 142829: gap of 100 bp  
\* 142830 167462: contig of 24633 bp in length.

FEATURES  
source

1. .167462  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-41418"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .1293  
/note="assembly\_fragment"

## misc\_feature

misc\_feature 1394..3228  
/note="assembly\_fragment"  
misc\_feature 3329..5140  
/note="assembly\_fragment"  
misc\_feature 5241..8017  
/note="assembly\_fragment"  
misc\_feature 8118..10466  
/note="assembly\_fragment"  
misc\_feature 10567..13763  
/note="assembly\_fragment"  
misc\_feature 13864..15896  
/note="assembly\_fragment"  
misc\_feature 15997..19987  
vector side:left  
misc\_feature 20088..24063  
/note="assembly\_fragment"  
misc\_feature 24164..26824  
/note="assembly\_fragment"  
misc\_feature 26925..30928  
vector side:right  
misc\_feature 31029..34504  
/note="assembly\_fragment"  
misc\_feature 34605..40409  
/note="assembly\_fragment"  
misc\_feature 40510..46642  
/note="assembly\_fragment"  
misc\_feature 46743..55317  
/note="assembly\_fragment"  
misc\_feature 55418..64038  
/note="assembly\_fragment"  
misc\_feature 64139..71781  
/note="assembly\_fragment"  
misc\_feature 71882..81517  
/note="assembly\_fragment"  
misc\_feature 81518..102057  
/note="assembly\_fragment"  
misc\_feature 102158..121599  
/note="assembly\_fragment"  
misc\_feature 121700..142729  
/note="assembly\_fragment"  
misc\_feature 142830..167462  
/note="assembly\_fragment"  
ORIGIN  
Query Match 68.5%; Score 290.4; DB 2; Length 167462;  
Best Local Similarity 86.3%; Pred. No. 1e-81;  
Matches 321; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
Qy 53 GGAGCTCTTTCTATTGCTTTCTCTCTGTCGCCATGCAGCTCATTTGGAAGGCAAAA 112  
Db 145485 GGAAGGTGTGTTAATGACTTCTACTTACTCTGTAATGCTGAATGTAATCTCTGAGTAA 145426  
Qy 113 AGGATAATCAGTTTCATCTGGAACACAGGTCGCGGGGAAGGTGTACAGGAGACTGTGGTC 172  
Db 145425 TTAACATCTTGTGTTTCTGCTGCTAGGTCCGTTGGGGAAGGTGTACAGGAGACTGTGGTC 145366  
Qy 173 CCGAGAGGTCAGAGTCGGGCGAGTGTGTTTTCATGTTGACGGTGGCAAGTCACC 232  
Db 145365 CCGAGAGGTCAGAGTCGGGCGAGTGTGTTTTCATGTTGACGGTGGCAAGTCACC 145306  
Qy 233 TGCTAACTGTGGTGAGAGCAACAGGCTCCAAAGGAAGAGTTGTTTCCGAGTTTGTG 292  
Db 145305 TGCTAACTGTGGTGAGAGCAACAGGCTCCAAAGGAAGAGTTGTTTCCGAGTTTGTG 145246  
Qy 293 ACTGGCAGTGACCTCTTTTCAGTGGGAGGTTTCTGACTGGCAGCACTGTGCTGTTTC 352  
Db 145245 ACTGGCAGTGACCTCTTTTCAGTGGGAGGTTTCTGACTGGCAGCACTGTGCTGTTTC 145186  
Qy 353 CTTACGCTCGGGTGAAGTCAAGCTCGGACTGCAGAGTGTGTGACGGCTCAGCATGGAC 412

|||||  
Db 145185 CTTACGCTCGCGTGAAGTCAAGCTCGACTGACAGAGTGTGTGACGGCTCAGATGGAC 145126  
|||||  
QY 413 TGCAGCACCGGA 424  
|||||  
Db 145125 TGCAGCACCGGA 145114  
|||||  
RESULT 4  
LOCUS AB051466 5724 bp mRNA linear PRI 07-FEB-2001  
DEFINITION Homo sapiens mRNA for KIAA1679 protein, partial cds.  
ACCESSION AB051466  
VERSION AB051466.1 GI:12697902  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K. and Ohara, O.  
Prediction of the coding sequences of unidentified human genes.  
XIX. The complete sequences of 100 new cDNA clones from brain which  
code for large proteins in vitro  
DNA Res. 7 (6), 347-355 (2000)  
21082932  
11214970  
2 (bases 1 to 5724)  
Ohara, O., Nagase, T. and Kikuno, R.  
Direct Submission  
Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Research Institute,  
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba  
292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp).  
URL: <http://www.kazusa.or.jp/huge>, Tel: 81-438-52-3913,  
Fax: 81-438-52-3914  
Location/Qualifiers  
1. 5724  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="fn22954"  
/note="vector:pBluescriptII SK plus"  
1. 5724  
/gene="KIAA1679"  
1. 4612  
/gene="KIAA1679"  
/note="Start codon is not identified."  
/codon\_start=2  
/product="KIAA1679 protein"  
/protein\_id="BAB21770.1"  
/db\_xref="GI:12697903"  
/translation="GWTSHLSNCGESNRPKRSKRCVCDWHSDFOWEVDWHHCVL  
VPYARGVKPTACVTAQHLQHRMVCIOKLNRTVVANEICEHFALOPPTQACLI  
PCPRDVSEFLPNSCKGGKLOHRTAVIAPLFGLOCPNLTESRACDAPISC  
PLGEEVTFSLKVPNSKCLPLHKEINPSGRTVLDPNSNSRNVTFKHSYXAHHS  
KSWAETGYQKVCSTNSQDAMLSLCLODSFPLVQSCIMPQCCTCSQSWSPFC  
SKTRSGSLPFGSRNRVGNMAIGGCKEPELLEKEACIVGELQQCPRYSWRTS  
EWKCVSLLEQDPHHVTGVPVCGGIQITREYVCAQSPAAAALAEKESVPVEKA  
LCVGPAPLPSQLNPGSTDCIVSSWAGWLCIHENQEPQKGFRTQRIHLMEST  
PGAGCPHLVSPCEPMDVLSSEGIQCPDHGKGLGHRILKAVCONDRGDEVSG  
SLCPVPPPERKSCIEICRMDVLSSEWTEMSQSCSNKNSDKOPRSRTILALAGE  
GKPCPSQALQHLRNLNDHSCMLHWTSPWGPCSDETLVATLNATIGNGATCGV  
GIQTRVVRKSHVQVTKKCPDSTRPETVPFCFLKDCDCTVATSEWTPCPRMCQ  
AGNATVQSRVRIIIOEAANGQCEPDTLYEERECEDVSLCPVVRKFPKWSPIVCP  
ESVAGTIGTSGEACGGLQTRAVSCIIDNRSKAEEMECLOKTMPLLVQETVPCRE  
DCTTANSGTFCCTGATRRRQLTGSKRKEKQDSDLPVLVTELCPCDFTIS  
QYVGNSDCIPEGRPHGLRVQADKECGELFRVACDKNGRPVDPSPFCSS  
GVIOBEKVIQCFPFCKLSDSWSSGSCSSGIGVIRISKKEPKYNGRCPEKLDLK  
NQAVHEAVPCYSCNQSWYVHWSCKINNELRSRCGGGTQSRKIRCVNTADGEG  
GAVSNLCADQEIPEPTQSCMLCPNCEWNLGWSKFCQSDPHTMQRTLRHLRP  
SLNRSKTCDSQVQFCLNENCFQYNLTWSTQCLSENAPCGQGVTRLLSCVSD  
GKPSVMDQCEQHLEKQKRMISIPCLVCEVNCQLSGMTAWTECSQTCGCGRMSRTF  
IIMPTQEGREPCPTLTQECTPTVPCYSWVLGNWSACKLEGDGCGEGVQIRSLSCW

HSGSISHAAGRVEDALCGEMPFPQDSILKQLCSVPCGDLTSEWSTCELTICIDGR  
SFETVGRSRRTPIIOSFENQDSCPOQVLETRPCTCKGKYHYTWKASLWNNERTYV  
CORSQGVNVTGGSPQAPPAIROCIIPACKRPFSYCTGGVCGEKEYTEIMRNGFL  
DYCKWPGSEDKADVKNLGKRNFKVNSKTHDIIPKWSLOPLDPDGRKVIWVYVSGG  
AFLIMIFLIFTSYLVCKKPRPHQSTPQKRELITLAYDGLDM"  
ORIGIN  
Query Match 49.3%; Score 209; DB 9; Length 5724;  
Best Local Similarity 100.0%; Pred. No. 1.2e-55;  
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 216 CGSGTGGACAGTACCTGTCTTAAGTGGTGAGAGCAACAGGCTCCAAAGGAAG 275  
Db 1 CGSGTGGACAGTACCTGTCTTAAGTGGTGAGAGCAACAGGCTCCAAAGGAAG 60  
QY 276 TTGTTTCCGAGTTTGTGACTGGCAGTGCCTCTTTTCAGTGGGAGGTTTCTGACTGGCA 335  
Db 61 TTGTTTCCGAGTTTGTGACTGGCAGTGCCTCTTTTCAGTGGGAGGTTTCTGACTGGCA 120  
QY 336 CCACGTGTGTCTTGTTCCTTACGTCGCGGTGAAGTCAAGCTCGGACTCGAGAGTGTGT 395  
Db 121 CCACGTGTGTCTTGTTCCTTACGTCGCGGTGAAGTCAAGCTCGGACTCGAGAGTGTGT 180  
QY 396 GACGCTCAGCATGACTGCAGCACCGGA 424  
Db 181 GACGCTCAGCATGACTGCAGCACCGGA 209  
RESULT 5  
AC111611 224789 bp DNA linear HTG 13-NOV-2002  
LOCUS Rattus norvegicus clone CH230-237p17, \*\*\* SEQUENCING IN PROGRESS  
DEFINITION \*\*\*  
AC111611  
AC111611.4 GI:24942115  
HTG; HTGS PHASE2; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 224789)  
Muzny, D. Marie., Metzger, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, R., Albrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Centz, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Eacotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gregg, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowals, C., Kratt, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorenshwari, L., Loulseghe, R., Lozado, R.J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Meenen, E.,  
Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, N., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwackeleme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Resier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.O., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tinney, A., Trejos, Z., Umanik, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 224789)  
Worley, K.C.  
Direct Submission  
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 224789)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 13, 2002 this sequence version replaced gi:22857292.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GMSK  
Center clone name: CH230-237P17  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 209106 bases at least Q40  
Consensus quality: 213906 bases at least Q30  
Consensus quality: 216632 bases at least Q20  
Estimated insert size: 218750; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter.  
\* This sequence will be replaced  
\* By the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 224789: contig of 224789 bp in length.

FEATURES  
source  
Location/Qualifiers  
1. 224789  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-237P17"

ORIGIN  
Query Match 47.3%; Score 200; DB 2; Length 224789;  
Best Local Similarity 80.9%; Pred. No. 1.3e-52;  
Matches 233; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 137 CAGTCGCTGGGAGAGTGTCACGAGAGAGTGTGTCGCGAGAGTCCAGAGTCGGGCGAG 196  
DB 43369 CAGGTCCTCGGGAAGGTGTGTCAGGTGACTGCGGACCGAGCGGAAACCCAGAGTCGCGCTG 43428  
QY 197 TGTGCTGTTTTTCATGTTGACGGGTGGACAGTACCTGTCTAACTGTGTTGGTGGAGCAACA 256  
DB 43429 TGTGCTGTTTTTCATGTTGACGGGTGGACAGTACCTGTCTAACTGTGTTGGTGGAGCAACA 43488  
QY 257 GGCTCCAAAGGAAAGTGTGTTGCGAGTGTGTCAGTGGCACAGTACCTCTTTTCAGT 316  
DB 43489 GACCTCCGAAGGAAGAGGCTGTTTCGCGTGTGTGACTGGCACAGTACCTCTTTTCAGT 43548  
QY 317 GGGAGGTTTCTGACTGCACACCTGTGTGCTTCTTACGCTCGGCTGAGTGAAGTCAAGC 376  
DB 43549 GGGAGGTTTCTGACTGCACACCTGTGTGCTTCTTACGCTCGGCTGAGTGAAGTCAAGC 43608  
QY 377 CTGCGACTGCAGAGTGTGTGACGGCTCAGCATGAGTGCAGCACCAGCA 424  
DB 43609 CTGCGGCTGTGGAGTGTGTGACTGCGCAGCAGCGTCTACAGCACCAGCA 43656

RESULT 6  
AC123741/c  
LOCUS  
DEFINITION Mus musculus clone RP24-113A5, WORKING DRAFT SEQUENCE, 5 unordered pieces.  
ACCESSION AC123741  
VERSION AC123741.4 GI:30017756  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 259236)  
Birren, B., Nusbaum, C. and Lander, E.  
Mus musculus, clone RP24-113A5  
Unpublished  
2 (bases 1 to 259236)  
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneu, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Vial, R., Vo, A., Wilson, M., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission

TITLE

## JOURNAL

Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 259236)

## REFERENCE

Authors: Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Canarata, J., Chang, J., Choepel, Y., Collumore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., McDermid, J., Menus, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Wu, X., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 17, 2003 this sequence version replaced gi:28882446.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

## TITLE

## JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WITBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information -----  
 Center project name: L26529  
 Center clone name: 113\_A\_5

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 655: contig of 655 bp in length  
 \* 656 755: gap of 100 bp  
 \* 756 1434: contig of 679 bp in length  
 \* 1435 1534: gap of 100 bp  
 \* 1535 7687: contig of 6153 bp in length  
 \* 7688 7787: gap of 100 bp  
 \* 7788 83305: contig of 75518 bp in length  
 \* 83306 83405: gap of 100 bp  
 \* 83406 259236: contig of 175831 bp in length.

## FEATURES

## source

1. .259236  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP24-113A5"  
 /clone\_lib="RPC1-24 Male Mouse BAC"  
 1. .655  
 /note="assembly\_fragment"  
 756. .1434  
 /note="assembly\_fragment"  
 1535. 7687  
 /note="assembly\_fragment"  
 7788. .83305  
 /note="assembly\_fragment"  
 83406. .259236  
 /note="assembly\_fragment"

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## ORIGIN

Query Match 46.0%; Score 195.2; DB 2; Length 259236;  
 Best Local Similarity 79.9%; Pred. No. 4.5e-51;  
 Matches 230; Conservative 0; Mismatches 58; Indels 0; Gaps 0;  
 QY 137 CAGGTCCCTGGGGAAGGTGTACAGGAGACTGTGTGTCGGGAGGAGTCCAGAGTCCGGCAG 196  
 Db 218756 CAGGTCCCTGGGGAAGGTGTACAGTGTGTCGACGACGAGCGGAGCCAAAGTCGTGCTG 218737  
 QY 137 TGTGGTGTTCATCTGTTTCACGGGTGGCAAGTCACTCTTAAGTGTGGTGAGAGCAACA 256  
 Db 218736 TATGGTGTCTTTTCATAGAGGGTGGACGAGTCCCATGTGCACTGTGATGAGAGCAGCC 218677  
 QY 257 GGCCTCCAAAGAAAGAGTGTTCGAGTTCCTGACTGGCAGTGCACAGTGACCTCTTCAGT 316  
 Db 218676 AACCTCCAAAGAAAGAGTGTTCGCGGTGTGACTGSCACAGTGACCTGTTCAGT 218617  
 QY 317 GGGAGGTTCCTGACTGGCACCACCTGTGTGCTTGTTCCTTACGCTCGCGGTGAAGTCAAGC 376  
 Db 218616 GGGAGGTTCCTGACTGGCACCACCTGTGTGCTTGTTCCTTCTGCTTCTGCGGCCCAAGGTGAGCCAGGC 218557  
 QY 377 CTCGGACTGCAGAGTGTGTGACGGCTCAGCATGACTGCAGCACCCGA 424  
 Db 218556 CTCGGCTGTGGAATGTGTGACTGCACAGCAGCGTCTACAGCACCCGA 218509

## RESULT 7

## AC079808

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 43 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1199: contig of 1199 bp in length
1200 1299: gap of unknown length
1300 2560: contig of 1261 bp in length
2561 2660: gap of unknown length
2661 4056: contig of 1396 bp in length
4057 4156: gap of unknown length
4157 5385: contig of 1229 bp in length
5386 5485: gap of unknown length
5486 7184: contig of 1698 bp in length
7185 7284: gap of unknown length
7285 9521: contig of 2238 bp in length
9522 11394: gap of unknown length
11395 11494: gap of unknown length
11495 13196: contig of 1702 bp in length
13197 13296: gap of unknown length
13297 14824: contig of 1528 bp in length
14825 14925: gap of unknown length
14926 16674: contig of 1750 bp in length
16675 16774: gap of unknown length
16775 18657: contig of 1883 bp in length
18658 18757: gap of unknown length
18758 20286: contig of 1528 bp in length
20287 20385: gap of unknown length
20386 22619: contig of 2234 bp in length
22620 22719: gap of unknown length
22720 25155: contig of 2436 bp in length
25156 25255: gap of unknown length
25256 27699: contig of 2444 bp in length
27700 27799: gap of unknown length
27800 29031: contig of 1232 bp in length
29032 29131: gap of unknown length
29132 31136: contig of 2005 bp in length
31137 31236: gap of unknown length
31237 33519: contig of 2283 bp in length
33520 33619: gap of unknown length
33620 35673: contig of 2054 bp in length
35674 35773: gap of unknown length
35774 37446: contig of 1673 bp in length
37447 37546: gap of unknown length
37547 40759: contig of 3213 bp in length
40760 40859: gap of unknown length
40860 43278: contig of 2419 bp in length
43279 43378: gap of unknown length
43379 46556: contig of 3178 bp in length
46557 46656: gap of unknown length
46657 50485: contig of 3829 bp in length
50486 50585: gap of unknown length
50586 53699: contig of 3074 bp in length
53700 53799: gap of unknown length
53800 57160: contig of 3401 bp in length
57161 57260: gap of unknown length
57261 60085: contig of 2805 bp in length
60086 60165: gap of unknown length
60166 64163: contig of 3998 bp in length
64164 64263: gap of unknown length
64264 67791: contig of 3528 bp in length
67792 67891: gap of unknown length
67892 72222: contig of 4331 bp in length
72223 72322: gap of unknown length
72323 76864: contig of 4542 bp in length
76865 76964: gap of unknown length
76965 81907: contig of 4943 bp in length
81908 82007: gap of unknown length
82008 87118: contig of 5111 bp in length

```

```

* 87119 87218: gap of unknown length
* 87219 92188: contig of 4970 bp in length
* 92189 92288: gap of unknown length
* 92289 97635: contig of 5347 bp in length
* 97636 97735: gap of unknown length
* 97736 104390: contig of 6655 bp in length
* 104391 104490: gap of unknown length
* 104491 110209: contig of 5719 bp in length
* 110210 110309: gap of unknown length
* 110310 117135: contig of 6826 bp in length
* 117136 117235: gap of unknown length
* 117236 123407: contig of 6172 bp in length
* 123408 123507: gap of unknown length
* 123508 130670: contig of 7163 bp in length
* 130671 130770: gap of unknown length
* 130771 132225: contig of 8455 bp in length
* 132226 139325: gap of unknown length
* 139326 150961: contig of 11636 bp in length
* 150962 151061: gap of unknown length
* 151062 164366: contig of 13305 bp in length.
FEATURES
    Location/Qualifiers
        1..164366
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="2"
            /clone="RP11-788B17"
        1..1139
            /note="assembly_name:Contig23"
        1300..2560
            /note="assembly_name:Contig26"
        2661..4056
            /note="assembly_name:Contig32"
        4157..5385
            /note="assembly_name:Contig33"
        5486..7183
            /note="assembly_name:Contig35"
        7284..9521
            /note="assembly_name:Contig36"
        9622..11394
            /note="assembly_name:Contig37"
        11495..13196
            /note="assembly_name:Contig38"
        13297..14824
            /note="assembly_name:Contig39"
        14925..16674
            /note="assembly_name:Contig40"
        16775..18657
            /note="assembly_name:Contig41"
        18758..20285
            /note="assembly_name:Contig42"
        20386..22619
            /note="assembly_name:Contig43"
        22720..25155
            /note="assembly_name:Contig44"
        25256..27699
            /note="assembly_name:Contig45"
        27800..29031
            /note="assembly_name:Contig46
            clone end:T7
            vector side:right"
        29132..31136
            /note="assembly_name:Contig47"
        31237..33519
            /note="assembly_name:Contig48"
        33620..35673
            /note="assembly_name:Contig49"
        35774..37446
            /note="assembly_name:Contig50"
        37547..40759
            /note="assembly_name:Contig51"
        40860..43278
            /note="assembly_name:Contig52"

```

```
misc_feature 43379..46556
              /note="assembly_name:Contig53"
misc_feature 46557..50485
              /note="assembly_name:Contig54"
misc_feature 50586..53659
              /note="assembly_name:Contig55"
misc_feature 53760..57160
              /note="assembly_name:Contig56"
misc_feature 57261..60065
              /note="assembly_name:Contig57"
misc_feature 60166..64163
              /note="assembly_name:Contig58"

Query Match      33.3%; Score 141; DB 2; Length 164366;
Best Local Similarity 100.0%; Pred.No. 1.2e-33;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGAGCAACTAACAGTCACTTCTGCTGGGTATGAGGAGCATGAGGAAGCTC 60
Db 77020 ATGTTTCCAAAGAGCAACTAACAGTCACTTCTGCTGGGTATGAGGAGCATGAGGAAGCTC 77079

QY 61 TTTCATTGCTTTCTCTCTCTGCTGCTCCATGCACTCATTTGGAAGCAAAAGGATAAT 120
Db 77080 TTTCATTGCTTTCTCTCTGCTGCTCCATGCACTCATTTGGAAGCAAAAGGATAAT 77139

QY 121 CAGTTCATCTGGAACACAGGT 141
Db 77140 CAGTTCATCTGGAACACAGGT 77160

RESULT 8
AC010873/c
LOCUS          180602 bp      DNA      linear      PRI 07-NOV-2001
DEFINITION    Homo sapiens BAC clone RP11-112N16 from 2, complete sequence.
ACCESSION     AC010873
VERSION       AC010873.12 GI:13992726
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 180602)
AUTHORS      Sulston,J.E. and Waterston,R.
TITLE        Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      99063792
PUBMED       9847074
REFERENCE     2 (bases 1 to 180602)
AUTHORS      Kyung,K. and Haakensen,W.
TITLE        The sequence of Homo sapiens BAC clone RP11-112N16
JOURNAL      Unpublished
REFERENCE     3 (bases 1 to 180602)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (25-SEP-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE     4 (bases 1 to 180602)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (08-MAY-2001) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
REFERENCE     5 (bases 1 to 180602)
AUTHORS      Waterston,R.
TITLE        Direct Submission
JOURNAL      Submitted (09-AUG-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE     6 (bases 1 to 180602)
AUTHORS      Waterston,R.
TITLE        Direct Submission
JOURNAL      Submitted (07-NOV-2001) Department of Genetics, Washington
              University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
```

## COMMENT

On May 8, 2001 this sequence version replaced gi:13518243.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0112N16  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-442L5, 200 bp overlap; the clone sequenced to the left is RP11-13K3, 200 bp overlap. Actual start of this clone is at base position 195 of RP11-112N16; actual end is at base position 180408 of RP11-112N16.

Polymorphic base differences occur in the overlap between RP11-112N16 and RP11-13K3.

The sequence from base position 151571 to 151695 and from position 151327 to 151337 is represented by a single strand as well as single chemistry.

## FEATURES

```
source
1..180602
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-112N16"
/clone_lib="RPCI-11"
repeat_region 1..73
/rpt_family="MIR"
repeat_region 255..500
/rpt_family="L2"
repeat_region 501..921
/rpt_family="MALR"
repeat_region 922..1317
/rpt_family="L2"
repeat_region 1331..2151
/rpt_family="L1"
repeat_region 2979..3101
/rpt_family="MIR"
repeat_region 3540..3870
```



```

/rpt_family="MALR"
3871..5452
/rpt_family="MALR"
5482..5901
/rpt_family="MALR"
6413..6703
/notes="similar to EST A1141965 (NID:g3649422) ool7e06.xl"
6413..6701
/notes="similar to EST AW962984 (NID:g8152820)"
6413..6628
/notes="similar to EST A1611205 (NID:g4620372) tw68h07.xl"
6417..6701
/notes="similar to EST A1830622 (NID:g5451293) wj10a04.xl"
6417..6695
/notes="similar to EST AA588635 (NID:g2402366) nj06d03.sl"
6417..6674
/notes="similar to EST A1090108 (NID:g3429167) qb28a11.xl"
6418..6703
/notes="similar to EST AW517645 (NID:g7155727) xt72c03.xl"
6425..6701
/notes="similar to EST AA358722 (NID:g2011041)"
6437..6703
/notes="similar to EST AA368086 (NID:g2020445)"
6437..6703
/notes="similar to EST AA902480 (NID:g3037603) ok70h12.sl"
6437..6703
/notes="similar to EST BF195923 (NID:g11083315)"
6454..6701
/notes="similar to EST AA358723 (NID:g2011042)"
6459..7064
/notes="similar to EST BE539677 (NID:g9768322)"
6459..6703
/notes="similar to EST AW383478 (NID:g6898241)"
6474..6704
/notes="similar to EST AW383489 (NID:g6898252)"
6479..6761
/notes="similar to EST BE739311 (NID:g10153389)"
6530..6759
/notes="similar to EST BF344750 (NID:g11292193)"
6530..6759
/notes="similar to EST BG163554 (NID:g12670257)"
6531..6755
/notes="similar to EST BE542005 (NID:g9770650)"
6556..6761
/notes="similar to EST BF086638 (NID:g10892257)"
6756..6807
/rpt_family="(TC)n"
6813..7064
/notes="similar to EST BF086638 (NID:g10892257)"
6813..7064
/notes="similar to EST BF344750 (NID:g11292193)"
6813..7064
/notes="similar to EST BG163554 (NID:g12670257)"
6813..7034
/notes="similar to EST BE739311 (NID:g10153389)"
6813..7005
/notes="similar to EST BE542005 (NID:g9770650)"
6873..7064
/notes="similar to EST H58372 (NID:g1011204) yr20b10.rl"
6880..7064
/notes="similar to EST AA469946 (NID:g2197255) zt98g04.sl"
7077..7449
/notes="similar to EST AW382106 (NID:g6898765)"
7077..7515
/notes="similar to EST BE568877 (NID:g9812597)"
7077..7436
/notes="similar to EST AW896456 (NID:g8060661)"
7135..7519
/notes="similar to EST BG036141 (NID:g12430992)"
7203..7442
/notes="similar to EST T81123 (NID:g704008) yd24a05.rl"
7211..7519
/notes="similar to EST BE393709 (NID:g9339061)"

```

```

misc_feature 7346..7521
/notes="similar to EST AW816566 (NID:g7909560)"
misc_feature 7349..7519
/notes="similar to EST AW361153 (NID:g6865903)"
misc_feature 7353..7519
/notes="similar to EST BG532258 (NID:g13523796)"
misc_feature 7420..7519
/notes="similar to EST AA278542 (NID:g1919927) zs77h10.rl"
misc_feature 7429..7519
/notes="similar to EST AW604121 (NID:g7308962)"
misc_feature 7440..7519
/notes="similar to EST BF247499 (NID:g11162992)"
misc_feature 7545..8028

Query Match 33.3%; Score 141; DB 9; Length 180602;
Best Local Similarity 100.0%; Pred. No. 1.2e-33;
Matches 141; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTTTCCAAGGACCACTTAACAGTCACCTGCTGGTATGGAGGAGCATGAGGAGCTC 60
Db 107961 ATGTTTCCAAGGACCACTTAACAGTCACCTGCTGGTATGGAGGAGCATGAGGAGCTC 107902
Qy 61 TTCTATTGCTTTCTCTCTCTGCTGCCATGCAGCTCATTGGAAGCAAAAAGGATAAT 120
Db 107901 TTCTATTGCTTTCTCTCTGCTGCCATGCAGCTCATTGGAAGCAAAAAGGATAAT 107842
Qy 121 CAGTTCATCTGGAACACGAGT 141
Db 107841 CAGTTCATCTGGAACACGAGT 107821

RESULT 9
AC096222/c
LOCUS AC096222 234528 bp DNA linear HTG 10-MAY-2003
DEFINITION Rattus norvegicus clone CH230-27C4, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC096222
VERSION AC096222.6 GI:30522362
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 234528)
Muzny, D., Marie, E., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Crease, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Deigado, L., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhewa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

```

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelenh, O., Okwundu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plummer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Reuter, M., Richards, S., Riggs, F., Rives, C., Rokey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, J., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Stearns, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wood, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 234528)  
Rat Genome Sequencing Consortium.  
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 10, 2003 this sequence version replaced gi:24942280.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GRM  
Center clone name: CH230-27C4  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 216647 bases at least Q40  
Consensus quality: 220163 bases at least Q30  
Consensus quality: 222500 bases at least Q20  
Estimated insert size: 232388; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.  
\* 1 234528: contig of 234528 bp in length.

# FEATURES

Location/Qualifiers  
1..234528  
/organism="Rattus norvegicus"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-27C4"

misc\_feature  
1..1930  
/note="wgs\_contig"

misc\_feature  
229739..231173  
/note="wgs\_contig"

misc\_feature  
232008..234528  
/note="wgs\_contig"

# ORIGIN

Query Match 26.1%; Score 110.6; DB 2; Length 234528;  
Best Local Similarity 86.5%; Pred. No. 7.3e-24;  
Matches 122; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGACGACCTACAGTCACCTGCTGCGGTATGAGGAGCATGAGGAGCTC 60

DB 151980 ATGTTTCCAAAGAGTGATCTGCGCTCACTCACTGGGTTTCGAGGAGCATGAGGAGCTC 151921

QY 61 TTCTATTGCTTTCTCTCTTCTGTCCTCCATGAGCTCATTTTGAAGCAAAAAGATAAT 120

DB 151920 TTCTAGTGTCTTCTCTCTGTCCTGTCCTCCCAAGCAGCTCATTTTGAAGGAGGAGGACAAAT 151861

QY 121 CAGTTCATCTGGAACACAGGT 141

DB 151860 CAGTTCATCTGGAACACAGGT 151840

# RESULT 10

AC124806/c

LOCUS

DEFINITION

AC124806

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC124806 175131 bp DNA linear HTG 16-DEC-2003  
Mus musculus chromosome 1 clone RP24-282F20 map 1, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 6 unordered pieces.

AC124806  
HTG: HTGS\_PHASE1: HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 175131)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus chromosome 1, clone RP24-282F20

Unpublished

2 (bases 1 to 175131)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,  
Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faroo, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D.,  
Galegari, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lakocque, K.,  
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,  
Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N.,  
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,  
Meneus, L., Mihov, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Pollara, V., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S.,  
Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 175131)

AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukagalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Menes, L., Mihova, T., Mlewa, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ranaasmy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission

JOURNAL Submitted (16-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Dec 16, 2003 this sequence version replaced gi:29294265. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L26619

Center clone name: 282\_F\_20

\* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 4834: contig of 4834 bp in length  
 4835 4934: gap of 100 bp  
 4935 63418: contig of 58484 bp in length  
 63419 63518: gap of 100 bp  
 63519 67337: contig of 3819 bp in length  
 67338 67437: gap of 100 bp  
 67438 147955: contig of 80518 bp in length  
 147956 148055: gap of 100 bp  
 148056 168460: contig of 20405 bp in length  
 168461 168560: gap of 100 bp  
 168561 175131: contig of 6571 bp in length.

## FEATURES

source

1..175131  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="1"  
 /map="1"  
 /clone="RP24-282F20"  
 /clone\_lib="RPCI-24 Male Mouse BAC"

## ORIGIN

Query Match 24.6%; Score 104.2; DB 2; Length 175131;  
 Best Local Similarity 83.7%; Pred. No. 8.1e-22;  
 Matches 118; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGACCACTAAACAGTCACCTGCTGGTATGGAGGACATGAGGAAGCTC 60  
 Db 76673 ATGTTTCTGAGGAGTGATCTGGCTGTCACTCACTGGGTTCCAGGAGCAAGGAGCTC 76614  
 QY 61 TTTCTATTGCTTTCTCTCTCTGCTGTGCTCCATCGAGCTCATTTGGAAGGCAAAAGATTAAT 120  
 Db 76613 TTTCTAGTGTCTTTCTCTCTGCTGTGCTCCCAAGCAGCTCATTTGGAAGGCAAGGACAAT 76554  
 QY 121 CAGTTTCATCTGGAACACAGGT 141  
 Db 76553 CAGTTTCCTCTGGAACACCGGT 76533

## RESULT 11

AX747433

LOCUS

AX747433

DEFINITION

Sequence 958 from Patent EP1308459.

ACCESSION

AX747433

VERSION

AX747433.1

GI:32131821

KEYWORDS

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J.I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tanechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and Masuho, Y.

TITLE

Full-length cDNA sequences

Patent: EP 1308459-A 958 07-MAY-2003;

Helix Research Institute (JP); Research Association for Biotechnology (JP)

JOURNAL

Location/Qualifiers

source

1..3053

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 18.5%; Score 78.6; DB 6; Length 3053;  
 Best Local Similarity 58.4%; Pred. No. 9.8e-14;  
 Matches 157; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

QY 90 TGCAGCTCATTTGGAAGGCAAAAGGATAATCAGTTTCATCTGGAACCCAGGTCCTCGTGGG 149  
 Db 393 TCCGCGCAGGCGGAGGCGGCCCCACCTCTATCTGTGGAAGACTGCTCATCGGG 452  
 QY 150 AAGGTGTACAGGA---GACTGTGTGTCCTCCGAGGAGTCCAGAGTGGGCGAGTGTGTTT 206  
 Db 453 CCGATGTATGCGAGATGATGTGTCCTCCGAGGAGTCCAAACGAGGCGCTGTGTGTGTC 512  
 QY 207 TCATGTTGACGGGTGGACAAGTCACTGTCTACTGTGTGAGAGCAACAGGCTCCAAA 266  
 Db 513 TCATGTTGAGGGATGGAGTCACTGATGATGATGATGATGATGATGATGATGATGAT 572  
 QY 267 GGAAGAAGATTGTTTCCGAGTTTGTGACTGGCAGTCAGTCACTCTTTTCAGTGGGAGTTTC 326  
 Db 573 CCAGCAGATTGTTTCAAAGTTTCCGATTGCGACAAAGAGTTGTACCTGGAGACTGGG 632  
 QY 327 TCATGTCACCACTGTGTGTTTTCCTTT 355  
 Db 633 ACCTTGAATCGGTGTCAGCCCGTGATT 661

## RESULT 12

AK092252

LOCUS

AK092252

DEFINITION

Homo sapiens cDNA FLJ34933 fis, clone NT2RP7005669.

ACCESSION

AK092252

VERSION

AK092252.1

GI:21750796

KEYWORDS

oligo capping; fis (full insert sequence).

SOURCE

Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,  
Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,  
Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,  
Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,  
Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y.,  
Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y.,  
Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A.,  
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,  
Nagatani, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project  
Unpublished

TITLE NEDO human cDNA sequencing project

JOURNAL Direct Submission

REFERENCE 2 (bases 1 to 3053)  
AUTHORS Isogai, T. and Yamamoto, J.  
TITLE Direct Submission  
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.  
Location/Qualifiers  
1...3053  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP7005669"  
/cell\_line="NT2"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2RP7"  
/notes="Cloning vector: pME18SPL3-mRNA from NT2 neuronal  
precursor cells after 5-weeks retinoic acid (RA)  
induction."

ORIGIN

Query Match 18.5%; Score 78.6; DB 9; Length 3053;  
Best Local Similarity 58.4%; Pred. No. 9.8e-14;  
Matches 157; Conservative 0; Mismatches 109; Indels 3; Gaps 1;

QY 90 TGCAGCTCATTTGGAAGGCAAAAGGATATACGTTTCATCTGGAAACCAGGTCCTGGGG 149  
Db 393 TCGCGCGAGGCGGAGCGGCGGCCCCCTCTATCTGTGAAGACTGTCCTCATGGGG 452

QY 150 AAGGTGTACAGGA---GACTGTGTCTCCGAGGAGTCCAGAGTCGGGCGAGTGTGGTGT 206  
Db 453 CCGATGTATGGGAGATGAATGTGTCTCCGAGGAGCATCAAACGAGGCTGTGGTGTGC 512

QY 207 TCATGTTCAGGGTGGACAGTCACTGTCTTAAGTGTGTGTGTGTGTGTGTGTGTGTGT 266  
Db 513 TCATGTGTGAGGATGGACTGACATGCACTGCACTGCACTGCACTGCACTGCACTGCA 572

QY 267 GGAAGAAGTGTGTTCGAGTTTGTGACTGGCAGTGACTCTTTTCAGTGGGAGGTTTC 326  
Db 573 CCAGCAGATTGTTTCAAGTTTTCGATTGGCACAAGAGTTGTACAGCTGGAGACTGGG 632

QY 327 TGACTGGCAGCAGT 355  
Db 633 ACCTTGAATCGGTGTGAGCCCGCTGATT 661

RESULT 13  
AR338841  
LOCUS AR338841 6378 bp DNA linear PAT 17-AUG-2003  
DEFINITION Sequence 332 from patent US 6569662.  
ACCESSION AR338841

AR338841.1 GI:33725698  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 6378)  
AUTHORS Tang, Y.T., Zhou, P. and Drmanac, R.T.  
TITLE Nucleic acids and polypeptides  
JOURNAL Patent: US 6569662-A 332 27-MAY-2003;  
FEATURES Location/Qualifiers  
1..6378  
source /organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Query Match 17.9%; Score 76; DB 6; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 7.2e-13;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;

QY 115 GATATCATGTTTCATCTGGAAACCAGGTCCTGGGGAGGAGTGTACAGGA---GACTGTGT 171  
Db 229 GGTAAATCGGAAAGCTTCTGAGAAAGTCCATGGGCGCGATGTATGGAGATGAATGTGT 288

QY 172 CCGGAGGAGTCCAGAGTCGGGCGAGTGTGGTGTTCATGTTCACGGGTGGACAAAGTCAC 231  
Db 289 CCGGAGGAGTCCAAACGAGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 348

QY 232 CTGTCTAAGT 291  
Db 349 CATACTAACTGTAAAGCAGCGCGAGAGACCCCAATAACAGCAGCAATTCGTTTCAAAAGTTTC 408

QY 292 GACTGGCACAGTGCCTCTTTTCAGTGGGAGGTTTCTGACTGGCACCAGTGTGTCTGTGT 351  
Db 409 GATTGGCACAAAGAGT 468

QY 352 CCTT 355  
Db 469 ATTT 472

RESULT 14  
AR338841  
LOCUS AR338841 6373 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 1 from Patent WO0105971.  
ACCESSION AR338841  
VERSION AR338841.1 GI:13159395  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1  
AUTHORS Shimkets, R.A. and Fernandes, E.  
TITLE Secreted polypeptides and corresponding polynucleotides  
JOURNAL Patent: WO 0105971-A 1 25-JAN-2001;  
FEATURES Location/Qualifiers  
1..6373  
source /organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
235..5001  
/note="unnamed protein product"

CDS  
1  
/codon\_start=1  
/protein\_id="CAC32422.1"  
/db\_xref="GI:13159396"  
/translation="MGDECGPGGIOTBAVNCVAHVCGHTTLHTNCKOARPNNOCNCFK  
VCDWHKELYDNRGLPWNQCPVSKSELEKICKEGQVREIACIQKDKIPIAD  
IICSYFEPKPLLEOALPCQDCIVSEFSAWSECKTCSGSLQHRTRHVVPFQGG  
SGCNLTFFQVCOSSCEBELRYSLVHWPSTCSMPHSQVQRARRGNKREKDR  
SKGVDPARLELIKKNRNQRQENKYWDIQIQTREVMCINKTKAADLSFCQQ  
EKLPMTFQSCVITKECVSEWSEWSPGSKTCHDMVSPAGTRVTRTRTIRPPIGSEKE



```
repeat_region /rpt_family="MIR"
repeat_region 7807.7846
repeat_region /rpt_family=" (TCRA)n"
repeat_region 7892.8181
repeat_region /rpt_family="Alu"
repeat_region 9128.9156
repeat_region /rpt_family="AT_rich"
repeat_region 9263.9974
repeat_region /rpt_family="L1"
repeat_region 10035.10061
repeat_region /rpt_family=" (CA)n"
repeat_region 10062.10941
repeat_region /rpt_family="L1"
repeat_region 10934.11278
repeat_region /rpt_family="L1"
repeat_region 11279.12784
repeat_region /rpt_family="L1"
repeat_region 12786.13758
repeat_region /rpt_family="L1"
repeat_region 13764.14207
repeat_region /rpt_family="L1"
repeat_region 14208.14564
repeat_region /rpt_family="MaLR"
repeat_region 14565.16156
repeat_region /rpt_family="L1"
repeat_region 16362.16948
repeat_region /rpt_family="ERV1"
repeat_region 17800.17956
repeat_region /rpt_family="L1"
repeat_region 17960.18238
repeat_region /rpt_family="MaLR"
repeat_region 18239.22161
repeat_region /rpt_family="L1"
repeat_region 22591.23172
repeat_region /rpt_family="MaLR"
repeat_region 23327.23381
repeat_region /rpt_family=" (TA)n"
repeat_region 23449.23510
repeat_region /rpt_family="GA-rich"
repeat_region 23586.24083
repeat_region /rpt_family="L1"
repeat_region 24084.24650
repeat_region /rpt_family="ERV1"
repeat_region 24651.24722
repeat_region /rpt_family="L1"
repeat_region 24735.24951
repeat_region /rpt_family="MIR"
repeat_region 25267.25287
repeat_region /rpt_family="AT_rich"
repeat_region 25317.25418
repeat_region /rpt_family="MIR"
repeat_region 25419.25705
repeat_region /rpt_family="Alu"
repeat_region 25706.25762
repeat_region /rpt_family="MIR"
repeat_region 26548.26594
repeat_region /rpt_family="AT_rich"
repeat_region 27187.27491
repeat_region /rpt_family="Alu"
repeat_region 29221.29497
repeat_region /rpt_family="L2"
repeat_region 30155.30315
repeat_region /rpt_family="MER1_type"
repeat_region 30941.31320
repeat_region /rpt_family="MaLR"
repeat_region 31385.31563
repeat_region /rpt_family="MER1_type"
repeat_region 32215.32484
repeat_region /rpt_family="Alu"
repeat_region 32485.32833
repeat_region /rpt_family="L1"
repeat_region 32841.33720
repeat_region /rpt_family="L1"
```

```
repeat_region 33721.34024
/rpt_family="Alu"
repeat_region 34025.34467
/rpt_family="L1"
repeat_region 35276.35443
/rpt_family="MIR"
repeat_region 35446.35919
/rpt_family="L1"
repeat_region 35920.36041
/rpt_family="Alu"
repeat_region 36120.36345
/rpt_family="L1"
repeat_region 36346.36613
/rpt_family="Alu"
repeat_region 36614.36827
/rpt_family="L1"
repeat_region 37155.37475
/rpt_family="L1"
repeat_region 37657.37711
/rpt_family="L2"
repeat_region 37884.37963
/rpt_family="L2"
repeat_region 38009.38226
/rpt_family="L2"
repeat_region 38267.38565
/rpt_family="Alu"
misc_feature 38295.38562
/note="CpG_island ($C=60.1, o/e=0.76, #CpGs=19)"

Query Match 17.8; Score 75.4; DB 9; Length 112456;
Best Local Similarity 62.0%; Pred. No. 1.4e-12;
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

QY 138 AGGTCCGTGGGGAGGTGTACAGGA---GACTGTGGTCCCGGAGGATCCAGAGTCGGGC 194
Db 100793 AGTCCATGGGGCCGATGTATGGGAGATGAATGTGTCGGGAGCATCCAAACGAGGCG 100734

QY 195 AGTGTGGTGTGTTTTCATGTTGACGGGTGGACAAAGTCACCTGTCTAACTGTGGTGAGAGCAA 254
Db 100733 TGTGTGGTGTGCTCATGTGGAGGGATGGACTACACTGCATACTAACTGTAAGCAGGCGA 100674

QY 255 CAGGCTCCAAAGGAAAGAGTGTGTTCCGAGTTGTGACTGGCAGACAGTGCACCTTTCA 314
Db 100673 GAGACCCAATAACACGAGCAATGTTTCAAAGTTTGGGATTGGCACAAGAGTTGTACGA 100614

QY 315 GTGGGAGGTGTTCTGACTGGCACCACCTGTGTGCTGTGTTCCIT 355
Db 100613 CTGGAGACTGGGACCTTGGATCAGTGTACGCCGTGATTT 100573
```

Search completed: February 23, 2004, 17:26:07  
Job time : 1598.77 secs

Result No.	Score	Query		DB	ID	Description
		Match	length			
1	286.8	67.6	5847	9	ADE071142	Novel
2	285	67.2	744	4	AAK34667	Human bon
3	285	67.2	744	4	AAKL7814	Human bon
4	284.4	62.4	477	4	AAK30658	Human bra
5	284.4	62.4	477	4	AAK05095	Human bon
6	78.6	18.5	3053	3	ADB62804	Human cDN
7	76	17.9	2326	3	AAC76705	Human ORF
8	76	17.9	6378	4	AAI58451	Human pol
9	76	17.9	6378	8	ADB48422	Novel hum
10	75.4	17.8	1956	4	AAI34372	Probe #30
11	75.4	17.8	1956	4	AAS28046	Human liv
12	75.4	17.8	6373	4	AAF30188	Clone 214
13	75.4	17.8	6373	7	ACC72162	Human NOV
14	75.4	17.8	6373	7	ACC72171	Human NOV
15	75.4	17.8	6373	7	ACC72168	Human NOV
16	74.8	17.6	6317	6	AAC38694	Human LP2
17	73.8	17.4	833	4	AAI47548	Probe #16
18	73.8	17.4	833	4	ABS41098	Human liv
19	73.8	17.4	921	7	ACC72173	Human NOV
20	73.8	17.4	4647	7	ACC72167	Human NOV
21	73.8	17.4	4679	7	ACC72166	Human NOV
22	73.8	17.4	4810	7	ACC72174	Human NOV
23	73.8	17.4	4881	7	ACC72165	Human NOV

CC proteins. The DNA and protein sequences of the invention are useful as:  
CC markers for tissues in which the corresponding protein is preferentially  
CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present DNA sequence represents a gene of the  
CC invention.

XX  
SQ Sequence 5847 BP; 1607 A; 1282 C; 1453 G; 1505 T; 0 U; 0 Other;  
Query Match 67.6%; Score 286.8; DB 9; Length 5847;  
Best Local Similarity 99.3%; Pred. No. 1.6e-81;  
Matches 288; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 135 ACCAGGTCGCGGGAAGGTGTACAGGAGACTGTGGTCCCGGAGGAGTCCAGAGTCGGGC 194  
Db 33 ACATGTCGCGTGGGGAAGGTGTACAGGAGACTGTGGTCCCGGAGGAGTCCAGAGTCGGGC 92  
QY 195 AGTCGGGTGTTTCATGTTGACGGGTGGACAAGTCACCTGCTAACTGTGTTGAGAGCAA 254  
Db 93 AGTGTGGGTGTTTCATGTTGACGGGTGGACAAGTCACCTGCTAACTGTGTTGAGAGCAA 152  
QY 255 CAGGCTCCAAAGGAAGAAGTGTGTTCCGAGTTTGTGACTGGCACAGTCACTCTTTTCA 314  
Db 153 CAGGCTCCAAAGGAAGAAGTGTGTTCCGAGTTTGTGACTGGCACAGTCACTCTTTTCA 212  
QY 315 GTGGAGGTTTCTGACTGGGACCACTGTGCTGCTTTCCTTACGCTCGCGTGAAGTCAA 374  
Db 213 GTGGAGGTTTCTGACTGGGACCACTGTGCTGCTTTCCTTACGCTCGCGTGAAGTCAA 272  
QY 375 GCTTCGAGTCCAGAGTGTGTGACGGGTGACATGAGTCCAGCACCGGA 424  
Db 273 GCTTCGAGTCCAGAGTGTGTGACGGGTGACATGAGTCCAGCACCGGA 322

RESULT 2  
AAK43667  
ID AAK43667 standard; DNA; 744 BP.  
XX  
AC AAK43667;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 18224.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000668.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024253.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PI WPI; 2001-488900/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.

PS Example 4; SEQ ID NO 18224; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention

XX  
SQ Sequence 744 BP; 181 A; 167 C; 201 G; 195 T; 0 U; 0 Other;  
Query Match 67.2%; Score 285; DB 4; Length 744;  
Best Local Similarity 100.0%; Pred. No. 2.5e-81;  
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 140 GTCCGTGGGGAAGGTGTACAGGAGACTGTGGTCCCGGAGGAGTCCAGAGTCGGCAGTGT 199  
Db 1 GTCCGTGGGGAAGGTGTACAGGAGACTGTGGTCCCGGAGGAGTCCAGAGTCGGCAGTGT 60  
QY 200 GGTGTTTCATGTTGACGGGTGGACAAGTCACTGTCTAACTGTGTTGAGAGCAACAGGC 259  
Db 61 GGTGTTTCATGTTGACGGGTGGACAAGTCACTGTCTAACTGTGTTGAGAGCAACAGGC 120  
QY 260 CTCCTAAAGCAAGAAAGTGTGTTCCGAGTTTGTGACTGGCACAGTCACTCTTTTCA 319  
Db 121 CTCCTAAAGCAAGAAAGTGTGTTCCGAGTTTGTGACTGGCACAGTCACTCTTTTCA 180  
QY 320 AGTGTTCGACTGGGACCACTGTGCTGCTTTCCTTACGCTCGCGTGAAGTCAAGCTC 379  
Db 181 AGTGTTCGACTGGGACCACTGTGCTGCTTTCCTTACGCTCGCGTGAAGTCAAGCTC 240  
QY 380 GACTGTCAGAGTGTGTGACGGGTGACATGAGTCCAGCACCGGA 424  
Db 241 GACTGTCAGAGTGTGTGACGGGTGACATGAGTCCAGCACCGGA 285

RESULT 3  
AAK17814  
ID AAK17814 standard; DNA; 744 BP.  
XX  
AC AAK17814;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 17805.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
XX ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000667.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
PI WPI; 2001-483446/52.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human



```
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX Example 4; SEQ ID NO 5215; 558pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX Q Sequence 744 BP; 181 A; 167 C; 201 G; 195 T; 0 U; 0 Other;
XX
XX Query Match 67.2%; Score 285; DB 4; Length 744;
XX Best Local Similarity 100.0%; Pred. No. 2.5e-81;
XX Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 140 GTCCGTGGGAGGTTGACAGAGACTGTGTCGCGGAGGTCAGAGTCCGAGTCCGCGAGTGT 199
XX 1 GTCCGTGGGAGGTTGACAGAGACTGTGTCGCGGAGGTCAGAGTCCGCGAGTGT 60
XX
XX 200 GGTGTTTTCATGTGACGGGTGGACAAAGTCACTGTCTAACTGTGCTGAGAGCAACAGGC 259
XX 61 GGTGTTTTCATGTGACGGGTGGACAAAGTCACTGTCTAACTGTGCTGAGAGCAACAGGC 120
XX
XX 260 CTCCTAAGGAAAGTGTGTTTCCGAGTTGTGACTGGCAGTCACTCTTTCACTGGG 319
XX 121 CTCCTAAGGAAAGTGTGTTTCCGAGTTGTGACTGGCAGTCACTCTTTCACTGGG 180
XX
XX 320 AGTGTTCCTGACTGGCACCACCTGTGCTGTTGTTCCCTAGCTCGCGGTGAAGTCAAGCCTC 379
XX 181 AGTGTTCCTGACTGGCACCACCTGTGCTGTTGTTCCCTAGCTCGCGGTGAAGTCAAGCCTC 240
XX
XX 380 GGACTGCAGAGTGTGACGGTCTGAGTGTGACTGGCAGTCAAGCAGCAGCA 424
XX 241 GGACTGCAGAGTGTGACGGTCTGAGTGTGACTGGCAGTCAAGCAGCAGCA 285
XX
XX RESULT 4
XX AK30658
XX D AAK30658 standard; DNA; 477 BP.
XX
XX AK30658;
XX
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed single exon probe SEQ ID NO: 5215.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
XX WO200157276-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
```

```
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX Example 4; SEQ ID NO 5215; 558pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX SQ Sequence 477 BP; 112 A; 94 C; 122 G; 149 T; 0 U; 0 Other;
XX
XX Query Match 62.4%; Score 264.4; DB 4; Length 477;
XX Best Local Similarity 85.3%; Pred. No. 9e-75;
XX Matches 295; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
XX
XX 53 GGAAGCTCTTTCTATTGCTTTCTCTCTTCTGTCCTCCATGCAAGTCAATTTGGAAAGGCAAAA 112
XX 132 GGAAGTGTGTTAATTGACTTCTACTTACCTGTGAATTCCTGAATTAATCTCTGAGTAA 191
XX
XX 113 AGGATAATCAGTTTCATCTGGAACCAAGTCCGTCGCGGAGGTCACAGAGACTGTGGTC 172
XX 192 TTAACATCTCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 251
XX
XX 173 CCGAGAGAGTCCAGAGTCCGCGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 232
XX 252 CCGAGAGAGTCCAGAGTCCGCGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 311
XX
XX 233 TGCTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 292
XX 312 TGCTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 371
XX
XX 293 ACTGGCAGTCACTCTCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 352
XX 372 ACTGGCAGTCACTCTCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 431
XX
XX 353 CTTACGCTCCGCTGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 398
XX 432 CTTACGCTCCGCTGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 477
XX
XX RESULT 5
XX AAK05095
XX ID AAK05095 standard; DNA; 477 BP.
XX
XX AC AAK05095;
XX
XX DT 05-NOV-2001 (first entry)
XX
XX DE Human brain expressed single exon probe SEQ ID NO: 5086.
XX
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
XX ss.
XX
XX OS Homo sapiens.
XX
XX EN WO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
```





XX Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
XX  
PS Claim 1; SEQ ID NO 654; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AA157798-AA161369) and the  
CC encoded polypeptides (AA158642-AA162213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 6378 BP; 1823 A; 1420 C; 1619 G; 1515 T; 0 U; 1 Other;

Query Match 17.9%; Score 76; DB 4; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 1.7e-13;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;  
QY 115 GATATCATGTTTCATCTGGAACACAGGTCCTGGGGAAGGTGTACAGGA---GACTGTGGT 171  
Db 229 GGTAAATGGGAAAGCTTCTGAGAAAGTCCATGGGCCGATGTGGGAGATGAATGTGGT 288  
QY 172 CCCGAGGAGTCCAGAGTCGGGCGAGTGTGTTTTCATGTTGACGGGTGGACAAGTCAC 231  
Db 289 CCCGAGGAGTCCAAACGAGGCTGTGTGTGTCTCATGTGGAGGATGACTACACTG 348  
QY 232 CTGCTTAAGTGTGTGAGCAACAGGCTCCAAAGGAAGAGTTGTTTCCGAGTTGT 291  
Db 349 CATACTTAAGTGTGAGCAACAGGCTCCAAAGGAAGAGTTGTTTCCAAAGTTGC 408  
QY 292 GACTGGCAGCTGACCTCTTTTCAGTGGGAGGTTTCTGACTGGCACCCTGTGTGCTGTT 351  
Db 409 GATTGGCAAAAGAGTTGTACAGCTGGAGCTGGGACCTTGGGAATCAGTGTACGCCGTG 468  
QY 352 CCTT 355  
Db 469 ATTT 472

RESULT 9  
ADB48422  
ID ADB48422 standard; cDNA; 6378 BP.  
XX  
AC ADB48422;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Novel human cDNA SEQ ID NO 332.  
XX  
KW ss; cancer; neurodegenerative disease; human.  
XX  
OS Homo sapiens.  
XX  
FN US2003104529-A1.  
XX  
PD 05-JUN-2003.  
XX  
PF 04-JAN-2002; 2002US-00037270.  
XX  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 19-JUL-2000; 2000US-00620312.  
XX

PA (ZHOU/) ZHOU P.  
PA (TANG/) TANG Y T.  
PA (LIUC/) LIU C.  
PA (ASUN/) ASUNDI V.  
PA (DRMA/) DRMANAC R T.  
XX  
PI Zhou P, Tang YT, Liu C, Asundi V, Drmanac RT;  
XX  
XX WPI; 2003-678194/64.  
DR  
XX New polynucleotide, useful for treating diseases e.g., cancer or  
PT neurodegenerative diseases.  
PT  
PS Claim 1; SEQ ID NO 332; 99pp; English.  
XX  
CC The invention relates to a polynucleotide comprising a sequence given in  
CC the specification, or its mature protein-coding portion, or its  
CC complement. The polynucleotide is useful for treating diseases e.g.,  
CC cancer or neurodegenerative diseases and many others listed in the  
CC specification. The present sequence represents a novel human cDNA. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from USPTO  
CC at [seqdata.uspto.gov/sequence.html?DocID=20030104529](http://seqdata.uspto.gov/sequence.html?DocID=20030104529).  
XX  
SQ Sequence 6378 BP; 1823 A; 1420 C; 1619 G; 1515 T; 0 U; 1 Other;

Query Match 17.9%; Score 76; DB 8; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 1.7e-13;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;  
QY 115 GATAATCATGTTTCATCTGGAACACAGGTCCTGGGGAAGGTGTACAGGA---GACTGTGGT 171  
Db 229 GGTAAATGGGAAAGCTTCTGAGAAAGTCCATGGGCCGATGTGGGAGATGAATGTGGT 288  
QY 172 CCCGAGGAGTCCAGAGTCGGGCGAGTGTGTTTTCATGTTGACGGGTGGACAAGTCAC 231  
Db 289 CCCGAGGAGTCCAAACGAGGCTGTGTGTGTCTCATGTGGAGGATGACTACACTG 348  
QY 232 CTGCTTAAGTGTGTGAGCAACAGGCTCCAAAGGAAGAGTTGTTTCCGAGTTGT 291  
Db 349 CATACTTAAGTGTGAGCAACAGGCTCCAAAGGAAGAGTTGTTTCCAAAGTTGC 408  
QY 292 GACTGGCAGCTGACCTCTTTTCAGTGGGAGGTTTCTGACTGGCACCCTGTGTGCTGTT 351  
Db 409 GATTGGCAAAAGAGTTGTACAGCTGGGACCTTGGGAATCAGTGTACGCCGTG 468  
QY 352 CCTT 355  
Db 469 ATTT 472

RESULT 10  
AAI34372  
ID AAI34372 standard; DNA; 1956 BP.  
XX  
AC AAI34372;  
XX  
DT 17-OCT-2001 (first entry)  
XX  
DE Probe #3058 used to measure gene expression in human placenta sample.  
XX  
KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW Genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
FN WO200157272-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US0000663.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
XX

PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00623366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488897/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 XX gene expression in human placenta.

PS Claim 25; SEQ ID NO 3058; 654pp; English.  
 CC  
 CC The present invention relates to single exon nucleic acid probes (SENP).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders  
 XX  
 SQ Sequence 1956 BP; 595 A; 370 C; 451 G; 540 T; 0 U; 0 Other;

Query Match 17.88; Score 75.4; DB 4; Length 1956;  
 Best Local Similarity 62.08; Pred. No. 1.6e-13;  
 Matches 13; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

QY 138 AGTCCCTGGGGAAGGTGTACAGGA---GACTGTGGTCCCGGAGGAGTCCAGAGTCGGGC 194  
 DB 189 AGTCCATGGGCGCGATGTATGGAGATGAATGTGTCGCGAGGATCCAAACGAGGC 248  
 2Y 195 ACGTGTGGTGTTCATGTGACGGGTGGACAGTCACTGTCTTAATGTGGTGAGAGCAA 254  
 DB 249 TGTGTGGTGTGCTCATGTGGAGGATGGACTACACTGCATCTAACTGTAAGCAGGCCGA 308  
 2Y 255 CAGGCCTCCAAAGAAAGATTGTTTCCGAGTTTGTGACTGGCAGAGTCACTCTTTCA 314  
 DB 309 GAGACCCATAACCCAGCAGATTGTTTCAAAGTTTGGATTGGCCAAAGAGTTGTACGA 368  
 2Y 315 GTGGAGGTTTCTGACTGGCACCAGTGTGCTGTTCCTT 355  
 DB 369 CTGGAGACTGGGACCTTGAATCAGTGTACGCCCGTGATTT 409

RESULT 11  
 ABS28046  
 ID ABS28046 standard; DNA; 1956 BP.  
 XX  
 AC ABS28046;  
 XX  
 DT 25-FEB-2003 (first entry)  
 XX  
 DE Human liver single exon probe, SEQ ID NO 3036.

XX Human; single exon nucleic acid probe; liver; cirrhosis;  
 KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
 KW coronary heart disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200157273-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000664.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00623366.

PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488898/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 XX gene expression in human adult liver.

PS Claim 1; SEQ ID NO 3036; 658pp; English.  
 CC  
 CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/fragments). The probe hybridizes at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABS25011-ABS51005 represent human  
 CC liver single exon nucleic acid probes of the invention. Note: The  
 CC sequence information for this patent does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1956 BP; 595 A; 370 C; 451 G; 540 T; 0 U; 0 Other;

Query Match 17.88; Score 75.4; DB 4; Length 1956;  
 Best Local Similarity 62.08; Pred. No. 1.6e-13;  
 Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
 QY 138 AGTCCCTGGGGAAGGTGTACAGGA---GACTGTGGTCCCGGAGGAGTCCAGAGTCGGGC 194  
 DB 189 AGTCCATGGGCGCGATGTATGGAGATGAATGTGTCGCGAGGATCCAAACGAGGC 248  
 QY 195 AGTGTGGTGTTCATGTGACGGGTGGACAGTCACTGTCTTAATGTGGTGAGAGCAA 254  
 DB 249 TGTGTGGTGTGCTCATGTGGAGGATGGACTACACTGCATCTAACTGTAAGCAGGCCGA 308  
 QY 255 CAGGCCTCCAAAGAAAGATTGTTTCCGAGTTTGTGACTGGCAGAGTCACTCTTTCA 314  
 DB 309 GAGACCCATAACCCAGCAGATTGTTTCAAAGTTTGGATTGGCCAAAGAGTTGTACGA 368  
 QY 315 GTGGAGGTTTCTGACTGGCACCAGTGTGCTGTTCCTT 355  
 DB 369 CTGGAGACTGGGACCTTGAATCAGTGTACGCCCGTGATTT 409

RESULT 12  
 AAF30188  
 ID AAF30188 standard; cDNA; 6373 BP.  
 XX  
 AC AAF30188;  
 XX  
 DT 30-APR-2001 (first entry)  
 XX  
 DE Clone 21433858 encoding secreted protein SECP1.  
 KW SECP1; secreted protein; human; diagnosis; gene therapy; lung cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 235..5001  
 FT sig\_peptide /\*tag= a  
 FT /\*tag= b  
 FT mat\_peptide 304..4998



Best Local Similarity 62.0%; Pred. No. 2.7e-13;  
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
138 AGGTCCGTGGGGAAGGTGTACAGGA---GACTGTGTGTCGCCGAGGAGTCCAGAGTCGGGC 194  
216 AGGTCCATGGGCGCGATGTATGGAGATGAATGTGTGTCGCCGAGGATCCAAACGAGGCG 275  
195 AGTGTGTTTTCATGTTGACGGGTGGCAAGTCACTGTCTAACTGTGTGGAGAGCAA 254  
276 TGTGTGTGTGTCTCATGTGGAGGATGGACTACATGCATACCTAACTGTAAAGCAGCGCA 335  
255 CAGGCTCCAAAGGAAGAGTGTTCGAGTTGTGACTGCGACAGTGCCTCTTTCA 314  
336 GAGACCCATAACCCAGCAGATTTGTTCAAGTTTCCGATTGCGCAAAAGAGTTGTACGA 395  
315 GTGGGAGGTTTGTGACTGCGACACACTGTGTGTGTTCTT 355  
396 CTGGAGACTGGGACCTTGGAAACAGTGTGAGCCCGTGATTT 436

## RESULT 14

ACC72171

:D ACC72171 standard; DNA; 6373 BP.

ACC72171;

07-JUL-2003 (first entry)

Human NOV47; coding sequence.

Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
immunomodulator; cytostatic; nootropic; neuroprotective;  
antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;  
diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;  
neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
immune disorder; haematopoietic disorder; gene; ds.

Homo sapiens.

WO2003029423-A2.

10-APR-2003.

02-OCT-2002; 2002WO-US031358.

02-OCT-2001; 2001US-0326483P.

05-OCT-2001; 2001US-0327342P.

09-OCT-2001; 2001US-0327917P.

09-OCT-2001; 2001US-0328029P.

09-OCT-2001; 2001US-0328044P.

09-OCT-2001; 2001US-0328056P.

12-OCT-2001; 2001US-0328849P.

15-OCT-2001; 2001US-0329414P.

17-OCT-2001; 2001US-0330142P.

22-OCT-2001; 2001US-0341058P.

24-OCT-2001; 2001US-0339266P.

24-OCT-2001; 2001US-0343629P.

29-OCT-2001; 2001US-0349575P.

01-NOV-2001; 2001US-0346357P.

12-APR-2002; 2002US-0371972P.

12-APR-2002; 2002US-0371980P.

17-APR-2002; 2002US-0373261P.

19-APR-2002; 2002US-0373805P.

23-APR-2002; 2002US-0374738P.  
16-MAY-2002; 2002US-0381101P.  
17-MAY-2002; 2002US-0381635P.  
29-MAY-2002; 2002US-0383830P.  
01-OCT-2002; 2002US-00262839.

(CURA-) CURAGEN CORP.

Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;

Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;  
Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
WPI; 2003-381625/36.  
P-PSDB; ABR58459.  
NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
dyslipidaemia, and in chromosome mapping, tissue typing or  
pharmacogenomics.

Claim 20; Page 263-264; 487pp; English.

The present invention relates to novel human NOV proteins and their  
coding sequences (ACC72075-ACC72181 and ABR58469). The NOV  
proteins are useful in manufacturing a medicament for treating a syndrome  
associated with a human disease. The NOV proteins and coding sequences  
may be used to diagnose, treat or prevent metabolic disorders such as  
diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
disorders such as Alzheimer's disease or Parkinson's disease, immune  
disorders, haematopoietic disorders and various dyslipidaemias

SQ Sequence 6373 BP; 1836 A; 1385 C; 1603 G; 1548 T; 0 U; 1 Other;

Query Match 17.8%; Score 75.4; DB 7; Length 6373;

Best Local Similarity 62.0%; Pred. No. 2.7e-13;

Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;

138 AGTCCGTGGGGAAGGTGTACAGGA---GACTGTGTGTCGCCGAGGAGTCCAGAGTCGGGC 194

216 AGTCCATGGGCGCGATGTATGGAGATGAATGTGTGTCGCCGAGGATCCAAACGAGGCG 275

195 AGTGTGTGTTTTCATGTTGACGGGTGGCAAGTCACTGTCTAACTGTGTGGAGAGCAA 254

276 TGTGTGTGTGTCTCATGTGGAGGATGGACTACACTGATCTAACTGTAAAGCAGCGCA 335

255 CAGGCTCCAAAGGAAGAGTGTTCGAGTTGTGACTGCGACAGTGCCTCTTTCA 314

336 GAGACCCATAACCCAGCAGATTTGTTCAAGTTTCCGATTGCGCAAAAGAGTTGTACGA 395

315 GTGGAGGTTTGTGACTGCGACACACTGTGTGTGTTCTT 355

396 CTGGAGACTGGGACCTTGGAAACAGTGTGAGCCCGTGATTT 436

## RESULT 15

ACC72168

:D ACC72168 standard; DNA; 6373 BP.

ACC72168;

07-JUL-2003 (first entry)

Human NOV47g coding sequence.

Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
immunomodulator; cytostatic; nootropic; neuroprotective;  
antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;  
diabetes; obesity; infection; cachexia; cancer; dyslipidaemia;  
neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
immune disorder; haematopoietic disorder; gene; ds.

Homo sapiens.

WO2003029423-A2.

10-APR-2003.

02-OCT-2002; 2002WO-US031358.

02-OCT-2001; 2001US-0326483P.

05-OCT-2001; 2001US-0327342P.

Search completed: February 23, 2004, 12:24:19



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: February 23, 2004, 11:46:56 ; Search time 968.64 Seconds  
(without alignments)  
13071.481 Million cell updates/sec

title: US-10-022-710-5\_COPY\_1\_424  
effect score: 424  
sequence: 1 agtttccaaagacaaact.....gcattggactgcagaccgga 424

coring table: IDENTIFY\_NUC  
Gapop 10.0 , Gapext 1.0

searched: 27513289 seqs, 14931090276 residues

total number of hits satisfying chosen parameters: 55026578

inimum DB seq length: 0  
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase :

EST:\*

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estrin.\*
- 4: em\_estmu.\*
- 5: em\_estov.\*
- 6: em\_estovl.\*
- 7: em\_estro.\*
- 8: em\_htc.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_hrc.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estom.\*
- 17: em\_gss\_hum.\*
- 18: em\_gss\_inv.\*
- 19: em\_gss\_pin.\*
- 20: em\_gss\_vrt.\*
- 21: em\_gss\_fun.\*
- 22: em\_gss\_mam.\*
- 23: em\_gss\_mus.\*
- 24: em\_gss\_pro.\*
- 25: em\_gss\_rcd.\*
- 26: em\_gss\_phg.\*
- 27: em\_gss\_vri.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

result	No.	Score	Match	Length	ID	Description
1	422.4	99.6	675	14	CF905312	CF905312 A0433A03-
2	294.4	69.4	626	10	BB655661	BB655661 BB655661
3	294.4	69.4	685	13	BU058391	BU058391 UI-M-FOO-
4	294.4	69.4	3242	11	AK083798	AK083798 Mus muscu

5	294.4	69.4	3897	11	AK051714	AK051714 Mus muscu
6	272	64.2	686	10	BB656175	BB656175 BB656175
7	213.6	50.4	474	10	BB656509	BB656509 BB656509
8	197.2	46.5	632	10	BB647268	BB647268 BB647268
9	196.6	46.4	2866	11	AK047508	AK047508 Mus muscu
10	171.6	40.5	411	13	BF706397	BF706397 280725 MA
11	119.8	24.6	456	10	AK052119	AK052119 Mus muscu
12	104.2	24.6	557	11	BB472930	BB472930 BB472930
13	104.2	24.6	565	10	BB658510	BB658510 BB658510
14	104.2	24.6	567	10	BB658510	BB658510 BB658510
15	102.2	24.1	676	14	CF531936	CF531936 UI-M-GHO-
16	94.4	22.3	704	14	CA361329	CA361329 635130 NC
c	79	18.6	983	29	CNS0368U	AL229827 Tetraodon
c	77.8	18.3	1201	9	AL531002	AL531002 AL531002
18	75.4	17.8	1148	11	AK080721	AK080721 Mus muscu
19	75.4	17.7	659	10	BB657304	BB657304 BB657304
20	75	17.7	1201	9	AL531003	AL531003 AL531003
21	67.6	15.9	620	10	BB656281	BB656281 BB656281
22	65.2	15.4	888	14	CB192844	CB192844 AGENCOURT
23	64.8	15.3	410	13	BY004406	BY004406 BY004406
24	62	14.6	392	13	BY004152	BY004152 BY004152
25	48.6	11.5	392	14	CK018416	CK018416 AGENCOURT
26	41	9.7	821	14	CK016846	CK016846 AGENCOURT
27	39.4	9.3	880	14	R83422	R83422 YP82f01.r1
28	39.2	9.2	433	14	BB728992	BB728992 BB728992
c	38.8	9.2	458	10	BG556716	BG556716 EM_40.F1
30	38	9.0	575	12	AW258602	AW258602 uq37c10.Y
31	37.4	8.8	463	10	CK015752	CK015752 AGENCOURT
32	36.6	8.6	869	14	CK015752	CK015752 AGENCOURT
33	36.4	8.6	478	10	BB284592	BB284592 BB284592
c	36.4	8.6	1101	29	CNS00KK2	AL077673 Drosophil
35	36.2	8.5	736	14	CB723795	CB723795 UI-M-FY0-
36	36.2	8.5	755	14	CB246260	CB246260 UI-M-FOO-
37	36.2	8.5	765	14	CA327230	CA327230 UI-M-FY0-
38	36	8.5	568	10	AW422427	AW422427 f162e12.Y
39	36	8.5	1201	13	BX446359	BX446359 BX446359
40	35.8	8.4	524	12	BG093822	BG093822 mab44f11.
41	35.8	8.4	532	14	CA890203	CA890203 B0157G12-
42	35.8	8.4	581	12	BM021611	BM021611 id42h07.X
43	35.8	8.4	584	12	BM021692	BM021692 id44b01.X
44	35.8	8.4	632	12	BI964985	BI964985 id34f08.X
45	35.8	8.4	650	14	CF539400	CF539400 UI-M-GHO-

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
CF905312	A0433A03-5	NIA Mouse Osteoblast cDNA Library (long 1) Mus musculus	CF905312	CF905312.1	GI:38172361	EST.	Mus musculus (house mouse)	1	Piao, Y., KO, N.T., Lim, M.K. and KO, M.S.H.	Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method	Genome Res. 11 (9), 1553-1558 (2001)	21429098	11544199	Contact: Dawood B. Dudekula Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@lsun.grc.nia.nih.gov Plate: AC433 row: A column: 03 Seq primer: M13 Reverse

High quality sequence stop: 675  
POLYA=No.

## FEATURES

## source

## Location/Qualifiers

1..675

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C3H/He mice"

/db\_xref="niaEST:A0433A03-5"

/db\_xref="taxon:10090"

/clone="NIA:A0433A03 IMAGE:30741122"

/dev\_stage="KUSA/A1 cells"

/lab\_host="DH10B"

/clone\_lib="NIA Mouse Osteoblast cDNA Library (Long 1)"

/notes="vector: pCMV-Sport6 (Invitrogen); Site 1: SalI;

Site 2: NotI; Mouse cDNA project by the Laboratory of

Genetics, National Institute on Aging (NIA), Intramural

Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).

This is a long-transcript enriched cDNA library (Ref.

Genome Res. 11: 1553-1558 (2001) [PMID: 11544199]). Total

RNAs were obtained from Dr. Akihiro Umezawa (Keio

University School of Medicine, Japan). Double-stranded

cDNAs were synthesized with an oligo(dT) primer

[Invitrogen]:

5'-pGATGCTTGTAGATCGGCGCGCCCTTTTTTTTTTTT-3' from

2.1 ug of total RNA, treated with T4 DNA polymerase, and

purified by ethanol-precipitation. The cDNAs were ligated

to Lona-linker LI-Sal4, purified by phenol/chloroform, and

separated from free linkers by Centricon 100. Then, the

cDNAs were amplified by long-range high fidelity PCR using

Ex Taq polymerase (Takara) with a primer Sal4-S. The

products were purified by phenol/chloroform and Centricon

100. The cDNAs were digested with SalI and NotI enzymes

and cloned into SalI/NotI site of pCMV-SPORT6 plasmid

vector. The DH10B E. coli host was transformed with the

ligation mixture by the standard chemical method. The

average insert size is about 3.0 kb. The library was

constructed by Yulan Piao."

## ORIGIN

Query Match

Best Local Similarity 99.6%; Score 422.4; DB 14; Length 675;

Matches 423; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGTTTCAAAGAGCAACCTTAACAGTCACCTTGCTGGTATGGAGGAGCATGAGGAAGCTC 60  
 DB 138 ATGTTTCAAAGAGCAACCTTAACAGTCACCTTGCTGGTATGGAGGAGCATGAGGAAGCTC 197  
 QY 61 TTTCTATTGCTTTCTCTGCTGCTCCATGCGAGCTCATTTGGAGGCAAAAGGATAAT 120  
 DB 198 TTTCTATTGCTTTCTCTGCTGCTCCATGCGAGCTCATTTGGAGGCAAAAGGATAAT 257  
 QY 121 CAGTTTCATCTGGAACACCAAGTCCGTGGGGAAGGTGTACAGGAGACTGTGTCCTCCGAGGA 180  
 DB 258 CAGTTTCATCTGGAACACCAAGTCCGTGGGGAAGGTGTACAGGAGACTGTGTCCTCCGAGGA 317  
 QY 181 GTCCAGAGTCGGGACAGTGTGGTGTTCATGTTGACGGGTGGCAAGTCACTGTCTTAAC 240  
 DB 318 GTCCAGAGTCGGGACAGTGTGGTGTTCATGTTGACGGGTGGCAAGTCACTGTCTTAAC 377  
 QY 241 TGTGCTGAGAGCAACAGGCTCCAAAGGAAGAGTGTTCCTCGAGTTTGTGACTGGCAC 300  
 DB 378 TGTGCTGAGAGCAACAGGCTCCAAAGGAAGAGTGTTCCTCGAGTTTGTGACTGGCAC 437  
 QY 301 AGTGACCTTTTCAGTGGAGGTTTTCAGTGGACCACTGTGCTGTTCCTTACGCT 360  
 DB 438 AGTGACCTTTTCAGTGGAGGTTTTCAGTGGACCACTGTGCTGTTCCTTACGCT 497  
 QY 361 CGCGGTGAAGTCAAGCTTCGAGTGCAGAGTGTGTGACGCTCAGCATGACATGCAGCAC 420  
 DB 498 CGCGGTGAAGTCAAGCTTCGAGTGCAGAGTGTGTGACGCTCAGCATGACATGCAGCAC 557  
 QY 421 CGGA 424  
 |||||

Db 558 CGGA 561

## RESULT 2

BB655661

LOCUS

DEFINITION

BB655661

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB655661 526 bp mRNA linear EST 26-OCT-2001  
 BB655661 RIKEN full-length enriched, 12 days embryo spinal ganglion  
 Mus musculus cDNA clone D13001L22 5', mRNA sequence.

BB655661

EST

BB655661.1 GI:16489489

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 526)

Arakawa, T., Carrincci, P., Fukuda, S., Furuno, M., Hanagaki, T.,

Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,

Kanno, H., Konda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,

Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,

Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.res.riken.go.jp/

Carrincci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Kanno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,

Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.

and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Kanno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,

Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamaraka, I.,

Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

e mouse tissues.

## FEATURES

## source

1..626

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="D13001L22"

/tissue\_type="spinal ganglion"

/dev\_stage="12 days embryo"

/lab\_host="DH10B"

/clone\_lib="RIKEN full-length enriched, 12 days embryo

spinal ganglion"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken



LOCUS  
DEFINITION  
AK083798 3242 bp mRNA linear HTC 20-SEP-2003  
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length  
enriched library, clone:DJ30011L22 product:weakly similar to  
HYPOTHETICAL PROTEIN DJ1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo  
sapiens], full insert sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AK083798.1 GI:26101501  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
99279253  
PUBMED  
10349636  
REFERENCE  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
20499374  
PUBMED  
11042159  
REFERENCE  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, K., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
PUBMED  
11076861  
REFERENCE  
4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
PANTOM Consortium  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 3242)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,  
URL: http://genome.res.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
7  
TAKEDA, Y., TANAKA, T., TOMARU, A., TOYA, T., YASUNISHI, A.,  
MURAMATSU, M. and HAYASHIZAKI, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,  
URL: http://genome.res.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
8  
ENCYCLOPEDIA PROJECT OF GENOME EXPLORATION RESEARCH GROUP IN RIKEN

Genomic Sciences Center and Genome Science Laboratory in RIKEN,  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.res.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers  
1. 3242  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM.DB:DJ30011L22"  
/db\_xref="MGI:2419369"  
/db\_xref="taxon:10090"  
/clone="DJ30011L22"  
/tissue type="spinal ganglion"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev stage="12 days embryo"  
1. 3242  
misc\_feature  
/note="weakly similar to HYPOTHETICAL PROTEIN DJ1110N13.1  
IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens] (SPTR:O43376,  
evidence: PASTY, 53.6%ID, 99.4%length, match=540)"  
ORIGIN  
Query Match 69.4%; Score 294.4; DB 11; Length 3242;  
Best Local Similarity 80.9%; Pred. No. 1.4e-80;  
Matches 343; Conservative 0; Mismatches 81; Indels 0; Gaps 0;  
QY 1 ATGTTTCAAAGAGCAACCTAACAGTCACTTCTGGGTATGAGGAGCATGAGGAAGCTC 60  
Db 160 ATGTTTCTGAGGAGTGATCTGGCTCACTCACTGGGTTTCCAGGAGCATGAGGAAGCTC 219  
QY 61 TTTCATTATGCTTTCTCTCTCTGCTGCCATCACTCATTTGGGAAGCAAAAGGATAAT 120  
Db 220 TTTCATTATGCTTTCTCTCTCTGCTGCCATCACTCATTTGGGAAGGAGGAGCAAT 279  
QY 121 CAGTTTCATCTCGAAACACAGGTCCTGGGGAGAGGTGTACAGGAGACTGTGGTCCCGAGGA 180  
Db 280 CAGTTTCCTCTGGAAACACCGTCCCTGGGGAGAGGTGTGCAGGTGACTGGGACCGAGGGA 339  
QY 181 GTCCAGAGTCGGGAGAGTGGTGTGTTCATGTCACGGTGGGACAGTCACTCTCTTAAC 240  
Db 340 GCCCAAGTCGTGTGTATGTGTCTTTCATAGAGGGTGGAGTCCCATGTCCGAAC 399  
QY 241 TGTGCTGAGAGCAACAGGCTCCCAAGAAAGAGTGTTCCTCCAGTTTGTGACTGGCAC 300  
Db 400 TGTGATGAGAGCAGCCACCTCCAAAGAGAGGAGCTGTTCCGCGTGTGTGACTGGCAC 459  
QY 301 AGTGACCTCTTTCAGTGGGAGTTTCTGACTGGGACCACTGTGTCTTGTCTTACGCT 360  
Db 460 AGTGACCTGTTTCCAGTGGGAGGTTTCCGACTGGCATCGCTGCTTCTGTTCTCTGGGGCC 519  
QY 361 CGCGGTGAAGTCAAGCCTCGGACTGCAGAGTGTGTGACGGTCAAGCATGAGTGCAGCAC 420  
Db 520 CAGGTGAGCCAGGCTCGGCTGTGGAATGTGTGACTGCACAGCAGCGTCTACAGCAC 579  
QY 421 CGGA 424  
Db 580 CGGA 583  
RESULT 5  
AK051714 3897 bp mRNA linear HTC 20-SEP-2003  
LOCUS  
DEFINITION  
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length  
enriched library, clone:DJ30067103 product:weakly similar to  
HYPOTHETICAL PROTEIN DJ1110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo  
sapiens], full insert sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AK051714  
AK051714.1 GI:26342139  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus



QY 421 CGGA 424

Db |||||

594 CGGA 597

## RESULT 6

BB656175

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB656175 686 bp mRNA linear EST 26-OCT-2001  
 BB656175 RIKEN full-length enriched, 12 days embryo spinal ganglion  
 Mus musculus cDNA clone D130067I03 5', mRNA sequence.

BB656175 GI:16490003

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 686)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,  
 Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
 Ohno, H., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
 Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,

Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.

and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,

Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I.,

Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and

Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

e mouse tissues.

Location/Qualifiers

1. 686

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="D130067I03"

/cissue\_type="spinal ganglion"

/dev\_stage="12 days embryo"

/lab\_host="DH10B"

/clone\_lib="RIKEN full-length enriched, 12 days embryo

spinal ganglion"

/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCAACTCGAGTGTGTTTTTTTTTTN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTTAATTAATTAATCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

## ORIGIN

Query Match 64.2%; Score 272; DB 10; Length 686;  
 Best Local Similarity 78.4%; Pred. No. 7, 4e-74;  
 Matches 323; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
 QY 1 ATGTTTCAAAGAGCAACTAACATCTGCTGGGTATGAGAGAGCATGAGGAAGCTC 60  
 Db |||||  
 174 ATGTTTCTGAGGAGTGATCTGCTGCTCACTCACTGGGTTTCAGGAGCATGAGGAAGCTC 233  
 QY 61 TTTCTATTGCTTTCTCTCTGCTGCCATCGAGTCAATTTGGAAGGCAAAAGGATAAT 120  
 Db |||||  
 234 TTTCTAGTCTTTCTCTCTGCTGCCAGAGCTCAATTTGGAAGGCAAAAGGATAAT 293  
 QY 121 CAGTTCATCTGGAACCAAGGTCCTGGGGAAGGTGTACAGGAGAGTGTGTCCCGGAGGA 180  
 Db |||||  
 294 CAGTTCCTCTGGAACCAAGGTCCTGGGGAAGGTGTGTGAGGTGAGTCTGGGAGCGGA 353  
 QY 181 GTCCAGAGTCCGGCAGTGTGTTTTCATGTTGACGGGTGAGCAAGTCACTGTCTTAAC 240  
 Db |||||  
 354 GCCCAAGTCGTGTATGTTGCTTTTCACTAGAAAGGTGAGCAGAGTCCCATGTCTGAAC 413  
 QY 241 TGTGTTGAGAGCAACAGGCTCCAAAGAAAGAAAGTGTGTTCCGAGTGTGTGACTGGCAC 300  
 Db |||||  
 414 TGTGATGAGAGCAGCAACCTCCAAAGGAAGGAGTGTGTTCCGGTGTGTGACTGGCAC 473  
 QY 301 AGTGACCTTTTCAGTGGAGGTTTCTGACTGGCAACCACTGTGTGCTTGTTCCTTACGCT 360  
 Db |||||  
 474 AGTGACCTGTTTCAGGGGAGGTTTCCCACTGGCATGCTGCTTTGTTGTTCTCTGGGNC 533  
 QY 361 CCGGTGAGTCAAGCCTCGGACTCGAGAGTGTGTGACGCTCAGCATGGAC 412  
 Db |||||  
 534 CCAGGGAGCCCGAGGCTTGGGCTGGGAGATGGNGACTGCACAGCCCGGCG 585

## RESULT 7

BB656509

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BB656509 474 bp mRNA linear EST 26-OCT-2001  
 BB656509 RIKEN full-length enriched, 12 days embryo spinal ganglion  
 Mus musculus cDNA clone D130033M03 5', mRNA sequence.

BB656509

BB656509.1 GI:16490337

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 474)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,

Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,

Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,

Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,

Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,

Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki



FEATURES  
source

## Location/Qualifiers

1. 632  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="B930082A18"  
 /sex="mixed"  
 /tissue\_type="cerebellum"  
 /dev\_stage="10 days neonate"  
 /lab\_host="DH10B"  
 /clone\_lib="RIKEN full-length enriched, 10 days neonate cerebellum"  
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGTCTCGAGTTAATTAAATTAATCCCTCCCCC Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"]

## ORIGIN

Query Match 45.5%; Score 197.2; DB 10; Length 632;  
 Best Local Similarity 80.0%; Pred. No. 2.1e-50;  
 Matches 232; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 135 ACCAGGTCCTGGGGAAGGTGTACAGGAGCTGTGGTCCCGAGGAGTCCAGAGTCGGGC 194  
 Db 247 AGCAGGTCCTGGGGAAGGTGTGACAGTGTGCGGACGAGCGGAGCCCAAGTCTGTC 306  
 Qy 195 AGTGTGGTGTTCATGTGACGGGTGACAGTCACTGTCTCACTGTGTGTGAGACCA 254  
 Db 307 TGTATGGTGTTCACATAGAGGAGGTGACAGTCCCATGTCGAACTGTGTATGAGACG 366  
 Qy 255 CAGGCTCCAAAGGAAGAAGTGTTCCTGAGTGTGTGACCTGGCACAGTGCACCTTTTCA 314  
 Db 367 CCAACTCCAAAGGAAGAAGTGTTCCTGAGTGTGTGACCTGGCACAGTGCACCTTTTCA 426  
 Qy 315 GTGGAGGTTTCTGACTGGCACTGTGTGCTGTTCCTTACGCTCGCGGTGAAGTCAA 374  
 Db 427 GTGGAGGTTTCCGACTGGCATCGTGTGCTTCTGCTTCTGGGCCCAAGGTGAGCCAG 486  
 Qy 375 GCTCGGACTCGAGAGTGTGTGACGGCTCAGCATGAGTGTGAGCAGCAGCA 424  
 Db 487 GCTCGGCTGGGAAATGTGTGACTGTGACACAGCGGTCTACAGCAGCA 536

## RESULT 9

AK047508  
 LOCUS 2266 bp mRNA linear HTC 20-SEP-2003  
 DEFINITION Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:B930082A18 product:weakly similar to HYPOTHETICAL PROTEIN DUIL10N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens], full insert sequence.

## ACCESSION

AK047508.1 GI:26092224

## VERSION

AK047508.1

## KEYWORDS

HTC; CAP trapper.

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

## REFERENCE

1

## AUTHORS

Carninci,P. and Hayashizaki,Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636

## REFERENCE

## AUTHORS

2  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159

## REFERENCE

## AUTHORS

3  
 Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Matakaki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 TITLE RIKEN integrated sequence analysis (KISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 PUBMED 11076861

## REFERENCE

## AUTHORS

4  
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 MEDLINE 11076861

## REFERENCE

## AUTHORS

5  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 MEDLINE 122666

## REFERENCE

## AUTHORS

6  
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Kanagaki,T., Kara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,O., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takada,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

## REFERENCE

## AUTHORS

7  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 MEDLINE 122666

## REFERENCE

## AUTHORS

8  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 MEDLINE 122666

## REFERENCE

## AUTHORS

9  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 MEDLINE 122666

## REFERENCE

## AUTHORS

10  
 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 MEDLINE 122666

## REFERENCE



```
/db_xref="taxon:10090"
/clone="B930082A18"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
misc_feature
1..2266
/note="weakly similar to HYPOTHETICAL PROTEIN D1110N13.1
IN CHROMOSOME 7 (FRAGMENT) [Homo sapiens] (SPTRI043376,
evidence: FASTY, 53.6%id, 99.4%length, match=540)"

>IGIN
Query Match 46.4%; Score 196.6; DB 11; Length 2266;
Best Local Similarity 79.7%; Pred. No. 66-50;
Matches 232; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

>Y 134 ACCAGGTCGGTGGGAAGGTGTACAGGAGATGTGGTCCCGAGGAGTCCAGATCGGG 193
>B 251 AGCCAGGTCCTCGGGAAGGTGTGCAGGTGACTGCGGACGAGCGGAGCCCAAGTCTGT 310
>Y 194 CAGTGTGGTGTTCATGTTCAGCGGTGACAGAGTCACTCTACTGTGTGTGAGACA 253
>B 311 CTGTATGGTCTTTCATAGAGGAGGTGACGAGTCCCATGTGCACTGTGTGAGAGCA 370
>Y 254 ACAGGCTCCAAAGGAAGAGTGTTCCTCGAGTGTGTGACTGGCACAGTCACTCTTTC 313
>B 371 GCCAACCCTCAAAGGAAGAGTGTTCCTCGGTGTGTGACTGGCACAGTCACTGTTC 430
>Y 314 AGTGGAGGTTTCTGACTGGCACCACTGTGTGTCTTACGCTCGCGTGAAGTCA 373
>B 431 AGTGGAGGTTTCCGACTGGCATCGTGTCTTCTGCTTCTCGGCGCCCAAGGTGAGCCA 490
>Y 374 AGCTCGGACTCGAGAGTGTGTGACGCTCAGCATGGAGTGGACGACCGGA 424
>B 491 GGCCTCGGCTGTGGATGTGTGACTGCACAGCAGGTCTACAGACCGGA 541

BY247109 411 bp mRNA linear EST 10-DEC-2002
BY247109 RIKEN full-length enriched, visual cortex Mus musculus
CDNA clone K230308M07 5', mRNA sequence.
BY247109
BY247109.1 GI:26428621
EST.
Mus musculus (house mouse)
Mus musculus
ORGANISM
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 411)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Oeato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojohori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Bruscia,V.,
Chochia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lehar,B., Lyons,P.A., Meglott,D.R.,
Maltais,L., Marchionni,L., McKenzia,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wyshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
```

Itch,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,  
Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,  
Rogers,J., Birney,E. and Hayashizaki,Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp/  
URL: http://genome.gsc.riken.go.jp/  
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,  
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,  
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,  
Onno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,  
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and  
Hayashizaki,Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Tissues were provided by Michela Pagiolini and Takao K. Hensch (  
Laboratory for Neuronal Circuit Development Brain Science Institute  
RIKEN 2-1 Hirosewa, Wako-shi, Saitama 351-0198 Japan) whose  
assistance we gratefully acknowledge. Please visit our web site  
(http://genome.gsc.riken.go.jp) for further details.

FEATURES  
Location/Qualifiers  
source

1..411  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K230308M07"  
/tissue\_type="visual cortex"  
/clone\_lib="RIKEN full-length enriched, visual cortex"

ORIGIN

Query Match 40.5%; Score 171.6; DB 13; Length 411;  
Best Local Similarity 81.8%; Pred. No. 1.9e-42;  
Matches 198; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGGACCACTTAACAGTCACCTTGTGGTATGGAGGAGCATGAGAGCTC 60  
DB 170 ATGTTTCTGAGGAGTGTATCGGCTGCTCACTCTGGGTTCCAGGAGCATGAGAGCTC 259  
QY 61 TTTCATTGCTTTCTCTCTTGTGTCCTCCATGTCAGCTCATTTGGAAGCAAAAGGATAT 120  
DB 230 TTTCATTGCTTTCTCTCTGCTGTCTCCAGCAGCTCATTTGGAAGCAGGAGGACAT 299  
QY 121 CAGTTCATCTGGAAAACAGGTCGCTGGGGAAGGTGTACAGGACCTGTGTGCCGAGGA 180  
DB 290 CAGTTCCTCTGGAAAACCGGTCCCTCGGGAAGGTGTGCAGGTGACTGCCGAGCGGA 349  
QY 181 GTCCAGAGTCGGGACGTGTGTGTCTTTTCATGTTGACGGTGGGACAAAGTCACCTGTCTAAC 240

```

Db 350 GCCCAAGTCGTGTATGGTCTTTTCACATAGAGGGTGGACGAGTCCCATGTGCGAAC 409
QY 241 TG 242
   ||
Db 410 TG 411

RESULT 11
LOCUS BF706397
DEFINITION 280725 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF706397
VERSION BF706397.1 GI:11998058
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
REFERENCE 1 (bases 1 to 456)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
        Casas,E., Wray,J.E., White,J., Cho,J., Fahrrenkrug,S.C.,
        Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
        Chitko-McKown,C.G., Pettea,G., Holt,I., Karamycheva,S., Liang,F.,
        Quackenbush,J. and Keefe,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
        libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
PUBMED 11282978
COMMENT Contact: Smith TPL
        USDA, ARS, US Meat Animal Research Center
        PO Box 166, Clay Center, NE 68933-0166, USA
        Tel: 402 762 4366
        Fax: 402 762 4390
        Email: smiththe@mail.marc.usda.gov
        Single pass sequencing. Bases called and alt trimmed with phred
        v0.980904.e. Vector identified by cross_match with the -minscore 18
        and -minmatch 12 options.
PCR PRIMERS
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCGC
Plate: 77 row: M column: 18
Seq primer: ATTTAGTGACACTATAG.
        Location/Qualifiers
FEATURES
     source
         1..456
         /organism="Bos taurus"
         /mol_type="mRNA"
         /db_xref="taxon:9913"
         /tissue_type="pooled"
         /lab_host="DH10B"
         /clone_lib="MARC 3BOV"
         /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
         Library made from pooled tissue from marrow, alveolar
         macrophage, ovary, fetal semitendinosus muscle, and fetal
         longissimus muscle."
ORIGIN
Query Match 28.3%; Score 119.8; DB 10; Length 456;
Best Local Similarity 91.4%; Pred No. 3.5e-26;
Matches 127; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATGTTTCCAAAGAGCAACTACAGTCACTGCTGGGTATCGAGAGCATGAGGAAGCTC 60
   |||||
Db 136 ATGTTTCCGAAGTAAATCTGGCTGCTCACTGCTGGGTATCGAGAGCATGAGGAAGCTA 195
   |||||
QY 61 TTTCATTGCTTTCCTCTGCTGTCCTCAGTCATTTGGAGGCAAAAGGATAAT 120
   |||||
Db 196 TTCTGTACITTCITCTCTTGCTGTGCTCCCATGAGCTCATTTGGAGGCAAAAGGATAAT 255
   |||||
QY 121 CAGTTTCATCTGGAAACCCAG 139
   |||||
Db 256 CAGTTTCATCTGGAAACCCAG 274

```

```

RESULT 12
LOCUS AK052119
DEFINITION
ACCESSION AK052119
VERSION AK052119.1 GI:26095006
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
        Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
        prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
        Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
        Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
        Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
        Fujisake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,S., Watahiki,M.,
        Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
        Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
        sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
        PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
        Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
        of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
        Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
        Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
        Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
        Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
        Koya,S., Kurikara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
        Nakamura,M., Nishi,K., Nomura,K., Numazaki,K., Ohno,M., Ohnato,N.,
        Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
        Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
        Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
        Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
        Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of
        Physical and Chemical Research (RIKEN), Laboratory for Genome
        Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
        RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
        Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,

```

```

AK052119 557 bp mRNA linear HTC 20-SEP-2003
Mus musculus 12 days embryo eyeball cDNA, RIKEN full-length
        enriched library, clone:D230048E01 product:weakly similar to
        HYPOTHETICAL PROTEIN DJ110N13.1 IN CHROMOSOME 7 (FRAGMENT) [Homo
        sapiens], full insert sequence.
ACCESSION AK052119
VERSION AK052119.1 GI:26095006
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

```

```

REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
        Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
        prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
        Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M.,
        Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
        Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
        Fujisake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,S., Watahiki,M.,
        Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
        Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
        sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
        PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
        Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
        of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
        Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
        Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
        Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
        Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
        Koya,S., Kurikara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
        Nakamura,M., Nishi,K., Nomura,K., Numazaki,K., Ohno,M., Ohnato,N.,
        Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
        Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
        Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
        Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
        Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of
        Physical and Chemical Research (RIKEN), Laboratory for Genome
        Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
        RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
        Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp,

```



```

QY 121 CAGTTCATCTGGAAACCAGGT 141
BB658510
Db 467 CAGTTCCTCTGGAAACCAGGT 487

RESULT 14
BB658510 567 bp mRNA linear EST 26-OCT-2001
LOCUS BB658510 RIKEN full-length enriched, 12 days embryo eyeball Mus
DEFINITION Musculus cDNA clone D330048E01 5', mRNA sequence.
ACCESSION BB658510
VERSION BB658510.1 GI:16492335
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 567)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,T.,
Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES             Location/Qualifiers
     source
     1..567
     /organism="Mus musculus"
     /mol_type="mRNA"
     /db_xref="taxon:10090"
     /clone="D330048E01"
     /tissue_type="eyeball"
     /dev_stage="12 days embryo"
     /lab_host="DH10B"
     /clone_lib="RIKEN full-length enriched, 12 days embryo
     eyeball"
     /note="Site_1: SalI; Site_2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGCGGCGCAACTCGAGTGTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5',
GAGAGAGAGATCTCGAGTTAAATAAATTAATCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN
Query Match 24.6%; Score 104.2; DB 10; Length 567;
Best Local Similarity 83.7%; Pred.No. 3.1e-21;
Matches 118; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 1 ATGTTTCCAAAGACCACTAACAGTCAGTCTGCTGGATGGAGGAGCATGAGGAAGCTC 60
Db 332 ATGTTTCTGAGGAGTGATCTGGCTGTCACTCACTGGGTTCAGGAGCATGAGGAAGCTC 391
QY 61 TTTCATATGCTTTTCTCTCTGCTGCTCCATGCAGCTCATTTTGGAAAGCAAAAAGGATAAT 120
Db 392 TTTCATGCTTTTCTCTCTGCTGCTCCAGCAGCTCATTTTGGAAAGCAGGAGCAAT 451
QY 121 CAGTTCATCTGGAAACCAGGT 141
Db 452 CAGTTCCTCTGGAAACCAGGT 472

RESULT 15
CF531936 676 bp mRNA linear EST 12-SEP-2003
LOCUS UI-M-GH0-CGU-0-15-0-UI.r1 NIH BMAP_GH0 Mus musculus cDNA clone
DEFINITION IMAGE:30356798 5', mRNA sequence.
ACCESSION CF531936
VERSION CF531936.1 GI:34583904
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uioawa.edu/distribution/mouseefi.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
     1..676
     /organism="Mus musculus"
     /mol_type="mRNA"
     /strain="C57BL/6"
     /db_xref="taxon:10090"
     /clone="IMAGE:30356798"
     /tissue_type="Whole brain"
     /dev_stage="1, 5, and 15 days newborn"
     /lab_host="DH10B (T1 phage resistant)"
     /clone_lib="NIH BMAP_GH0"
     /note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
     Site_2: Not I; The library was constructed according

```

Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGTGAAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., Program coordinator."

RIGIN

Query Match	24.1%	Score	102.2	DB	14	Length	676
Best Local Similarity	83.5%	Pred. No.	1.4e-20				
Matches	116	Conservative	0	Mismatches	23	Indels	0
		Gaps	0				
y	1	ATGTTTCCAAAGACCACTAACAGTCACTTGGTGGTATGGAGGAGCATGAGGAGCTC	60				
b	151	ATGTTTCTGAGAGGATGCTGCTCTCACTCACTGGGTTCCAGGAGCATGAGGAGCTC	210				
y	61	TTTCTATTGCTTTCTCTCTTGTGTCCTCCATGCGCTCATTTTGGAGGCAAAAGGATAAT	120				
b	211	TTTCTAGTGCCTTTCTCTCTCTGTCCTCCCAAGCAGCTCATTTTGGAGGAGGAGGACAAT	270				
y	121	CAGTTCATCTGGAACACAG	139				
b	271	CAGTTCCTCTGGAACACCG	289				

search completed: February 23, 2004, 20:29:37  
Ob time : 973.64 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 12:03:01 ; Search time 31.7428 Seconds  
(without alignments)  
7412.668 Million cell updates/sec

Title: US-10-022-710-5\_COPY\_1\_424

Perfect score: 424

Sequence: 1 atgtttccaaagacacct.....geatggactgcagcaccgga 424

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
  - 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
  - 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
  - 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
  - 5: /cgn2\_6/ptodata/2/ina/PCFUS COMB.seq.\*
  - 6: /cgn2\_6/ptodata/2/ina/backfileseq1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76	17.9	6378	4	US-09-620-312D-332
2	33.2	7.8	695	4	US-09-554-451-2
3	31	7.3	2866	4	US-09-724-864-30
4	30.8	7.3	1083	4	US-09-107-532A-869
5	30.6	7.2	239	4	US-09-642-703-33
6	30.6	7.2	288	4	US-09-642-703-34
7	30.6	7.2	1664976	4	US-08-916-421B-1
8	30.4	7.2	2824	4	US-07-757-022B-13
9	30.4	7.2	3066	4	US-07-757-022B-83
10	30.4	7.2	3117	4	US-07-757-022B-73
11	30.4	7.2	3148	4	US-07-757-022B-57
12	30.4	7.2	3420	4	US-07-757-022B-103
13	30.4	7.2	3813	4	US-07-757-022B-43
14	30.4	7.2	3936	4	US-07-757-022B-41
15	30.4	7.2	3942	4	US-07-757-022B-141
16	30.4	7.2	3945	4	US-07-757-022B-49
17	30.4	7.2	3963	4	US-07-757-022B-45
18	30.4	7.2	3963	4	US-07-757-022B-59
19	30.4	7.2	4065	4	US-07-757-022B-47
20	30.4	7.2	4086	4	US-07-757-022B-39
21	30.4	7.2	4092	4	US-07-757-022B-51
22	30.4	7.2	4215	4	US-07-757-022B-61
23	30.4	7.2	4847	4	US-10-164-595-57
24	30.4	7.2	5008	4	US-07-757-022B-1
25	30.4	7.2	5041	4	US-09-023-655-981
26	29.8	7.0	2342	4	US-09-620-312D-989
27	29.6	7.0	1022	3	US-09-189-035-4

Sequence 4, Appli  
Sequence 18, Appli  
Sequence 1349, Ap  
Sequence 14, Appli  
Sequence 1, Appli  
Sequence 7072, Ap  
Sequence 7027, Ap  
Sequence 5349, Ap  
Sequence 5192, Ap  
Sequence 6976, Ap  
Sequence 5283, Ap  
Sequence 327, App  
Sequence 14891, A  
Sequence 14740, A  
Sequence 15015, A  
Sequence 5, Appli  
Sequence 59, Appli  
Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-620-312D-332  
; Sequence 332, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yundong  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL genes Version 1.0  
; SEQ ID NO 332  
; LENGTH: 6378  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (271)..(5037)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(6378)  
; OTHER INFORMATION: n = a,t,c or g  
US-09-620-312D-332

Query Match 17.9%; Score 76; DB 4; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 3.7e-15;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;

QY 115 GATAATCAGTTTCATCTGGAACACAGGTCCTGGGGAAGGTGTACAGGA---GACTGTGGT 171  
Db 229 GGTAATTGGGAAGCTTCTGAGAAAGTCCATGGGCGGATGTATGGAGATGAATGTGGT 288  
QY 172 CCGGAGAGTCCAGAGTCGGGAGTGTGTTTTCATGTTACGGGTGGACAATGCAC 231  
Db 289 CCGGAGGATCCAAACGAGGGCTGTGTGTGTCTCATGTGGAGGATGATACACTG 348  
QY 232 CTGTCTAACTGTGTGAGACACAGCCCTCCAAAGGAAGAAAGTTGTTTCCAGTTGT 291  
Db 349 CATACTAATCTAAGCAGCGCGGAGACCCCAATPACCAGCAGATTGTTTCAAGTTGC 408  
QY 292 GACTGGCAGAGTACCTCTTTCAGTGGAGGTTTCTGACTGGACACACTGTGTGTTGT 351  
Db 409 GATTGGCACAAGAGTTGTACGACTGGAGACTGGACCTTGGAAATCAGTGTGACCGGTG 468  
QY 352 CCTT 355  
Db 469 ATTT 472

RESULT 2  
US-09-554-451-2/c  
; Sequence 2, Application US/09554451  
; Patent No. 6680207  
; GENERAL INFORMATION:  
; APPLICANT: Jonathan Paul MURPHY  
; TITLE OF INVENTION: Detection of Molecules in Samples  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pillsbury Winthrop, L.L.P.  
; STREET: 1100 New York Ave., N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB98/03449  
; FILING DATE: No. 6680207ember 16, 1998  
; APPLICATION NUMBER: GB 9723955.2  
; FILING DATE: No. 6680207ember 14, 1997  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 695 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-554-451-2

Query Match 7.8%; Score 33.2; DB 4; Length 695;  
Best Local Similarity 53.0%; Pred. No. 0.29; Mismatches 63; Indels 0; Gaps 0;  
Matches 71; Conservative 0;  
QY 136 CCAGGTCCGTGGGAGGTGTACAGAGACTGTGTCGGGAGGAGTCCAGAGTCGGCA 195  
Db 624 CCATGCTTTTACGGAAGCAGTACAGACCGTAGTTTTTCAGCAGAGCGTCGTGTTGT 565  
QY 196 GTGTGGTGTTCATGTTGACGGTGGACAGTACCTGTCTTAATGTGGTGGAGCAAC 255  
Db 564 GAGAGTGTGTGCAATTTAGATAGTGTCTGTTGAAGATCTGACCGGTACCGGAGAAC 505  
QY 256 AGGCCTCCAAAGA 269

Db 504 CGTCTTCCAGACGA 491  
RESULT 3  
US-09-724-864-30  
; Sequence 30, Application US/09724864  
; Patent No. 6380362  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Murison, James G.  
; TITLE OF INVENTION: Polynucleotides, polypeptides expressed  
; TITLE OF INVENTION: by the polynucleotides and methods for their use.  
; FILE REFERENCE: 11000.1050U1  
; CURRENT APPLICATION NUMBER: US/09/724,864  
; CURRENT FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 2866  
; TYPE: DNA  
; ORGANISM: Mouse  
US-09-724-864-30  
Query Match 7.3%; Score 31; DB 4; Length 2866;  
Best Local Similarity 50.3%; Pred. No. 3.3; Mismatches 75; Indels 0; Gaps 0;  
Matches 76; Conservative 0;  
QY 220 TGGACAAGTCACTGTCTTAACCTGTGTGAGAGCAACAGCGCTCCAAAGAAAGTTGT 279  
Db 592 TGGACGAATCTCTTGGCCCTCGAATGTGAATGGAATGCAATGAAATGAAGGAACAGG 651  
QY 280 TTCCGAGTTTGTGACTGGCAGACAGTCACTCTTTCAGTGGGAGTTTCTGACTGGCACCAC 339  
Db 652 GTCCAGCGGGGCAACACAGTTTACCAGTACCTCTTCCATGGTACCTTCTCCGACGC 711  
QY 340 TGTGTGCTTGTTCCTTAGCTCGCGTGAAG 370  
Db 712 AGACGGCAAGACATCACTGCACGTTGAG 742

RESULT 4  
US-09-107-532A-869  
; Sequence 869, Application US/09107532A  
; Patent No. 5583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A. Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:





```
, LOCATION: (84773)..(84773)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (622708)..(622708)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (657081)..(657081)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (657203)..(657203)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (674435)..(674435)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (682442)..(682442)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (713652)..(713652)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (741684)..(741684)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (779455)..(779455)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (779676)..(779676)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (855539)..(855539)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (871619)..(871619)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1084830)..(1084830)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1095846)..(1095846)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1119881)..(1119881)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1130881)..(1130881)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1310988)..(1310988)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1313224)..(1313224)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1345473)..(1345473)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1349491)..(1349491)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1470091)..(1470091)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1569020)..(1569020)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1602912)..(1602912)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1603734)..(1603734)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (1637998)..(1637998)
, OTHER INFORMATION: n equals a, t, c, or g
, LOCATION: (84773)..(84773)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (84808)..(84808)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (84812)..(84812)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98120)..(98120)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98159)..(98159)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98239)..(98239)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98266)..(98266)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (98343)..(98343)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (103998)..(103998)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (148948)..(148948)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (163385)..(163385)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (191989)..(191989)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (191995)..(191995)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (231980)..(231980)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (234187)..(234187)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (234220)..(234220)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (234814)..(234814)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (309398)..(309398)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (309418)..(309418)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (312837)..(312837)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (312993)..(312993)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (319226)..(319226)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (559167)..(559167)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (559241)..(559241)
, OTHER INFORMATION: n equals a, t, c, or g
, NAME/KEY: misc feature
, LOCATION: (600392)..(600392)
```





;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: double  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: CDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..3147  
JS-07-757-022B-57

Query Match 7.2%; Score 30.4; DB 4; Length 3148;  
Best Local Similarity 55.7%; Pred. No. 5.5;  
Matches 78; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
2Y 58 CTCCTTCTATTGCTTCTCTCTGCTGTCCTCCATGCGAGCTCATTGGAAGGCAAAAAGGAT 117  
2b 2494 CTCCTTCTGTTGCTTCTCTTTAGCTGTTGTAGTCATTTCAGCTTTAGTCGCTGCAGGAG 2435  
2Y 118 AATCAGTTCACTGGAACACAGGTCGCTGGGGAAGGTGTACAGGAGACTGTGTCCTCCGGA 177  
2b 2434 TCTTAGTT-GTAGGTACACAGGTTCTCTGGGACTGTTTTCAGAGCTTTTGGTGTGGGT 2376  
2Y 178 GGAGTCAGAGTCGGGCAGT 197  
2b 2375 TCTGCAGAAAGCTCAGGAGT 2356

RESULT 12  
US-07-757-022B-103/c  
; Sequence 103, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 103:

;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 3420 base pairs  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: double  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: CDNA  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..3420  
US-07-757-022B-103

Query Match 7.2%; Score 30.4; DB 4; Length 3420;  
Best Local Similarity 55.7%; Pred. No. 5.7;  
Matches 78; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
QY 58 CTCCTTCTATTGCTTCTCTCTGCTGTCCTCCATGCGAGCTCATTGGAAGGCAAAAAGGAT 117  
Db 2767 CTCCTTCTGTTGCTTCTCTTTAGCTGTTGTAGTCATTTCAGCTTTAGTCGCTGCAGGAG 2708  
QY 118 AATCAGTTCACTGGAACACAGGTCGCTGGGGAAGGTGTACAGGAGACTGTGTCCTCCGGA 177  
Db 2707 TCTTAGTT-GTAGGTACACAGGTTCTCTGGGACTGTTTTCAGAGCTTTTGGTGTGGGT 2649  
QY 178 GGAGTCAGAGTCGGGCAGT 197  
Db 2648 TCTGCAGAAAGCTCAGGAGT 2629

RESULT 13  
US-07-757-022B-43/c  
; Sequence 43, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3813 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3810  
US-07-757-022B-43

Query Match 7.2%; Score 30.4; DB 4; Length 3813;  
Best Local Similarity 55.7%; Pred. No. 6;  
Matches 78; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
QY 58 CTCCTTCTATGCTTCTCTCTGTCGCCATGCGAGCTCATTTGGAGGCAAAAGGAT 117  
Db 2365 CTCCTTCTATGCTTCTCTCTGTCGCCATGCGAGCTCATTTGGAGGCAAAAGGAT 117  
QY 118 AATCAGTTCACTGGAACCCAGGTCGGTGGGAAGGTGTACAGGAGCTGTGTCGCCGA 177  
Db 2305 TCTTAGTT-GTAGGTACACAGGTCCTTGGGACTGTTTCAAGAGCTTTTGGTGGGT 2247  
QY 178 GGAGTCCAGAGTCGGGCACT 197  
Db 2246 TCTGCAGAAAGCTCAGAGT 2227

RESULT 14  
US-07-757-022B-41/c  
Sequence 41, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3936 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3933  
US-07-757-022B-41

Query Match 7.2%; Score 30.4; DB 4; Length 3936;  
Best Local Similarity 55.7%; Pred. No. 6.1;  
Matches 78; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
QY 58 CTCCTTCTATGCTTCTCTCTGTCGCCATGCGAGCTCATTTGGAGGCAAAAGGAT 117  
Db 2488 CTCCTTCTATGCTTCTCTCTGTCGCCATGCGAGCTCATTTGGAGGCAAAAGGAT 117  
QY 118 AATCAGTTCACTGGAACCCAGGTCGGTGGGAAGGTGTACAGGAGCTGTGTCGCCGA 177  
Db 2428 TCTTAGTT-GTAGGTACACAGGTCCTTGGGACTGTTTCAAGAGCTTTTGGTGGGT 2370  
QY 178 GGAGTCCAGAGTCGGGCACT 197  
Db 2369 TCTGCAGAAAGCTCAGAGT 2350

RESULT 15  
US-07-757-022B-141/c  
Sequence 141, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann

REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 141:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3942 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3939  
US-07-757-022B-141

Query Match 7.2%; Score 30.4; DB 4; Length 3942;  
Best Local Similarity 55.7%; Pred.No.6.1;  
Matches 78; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
Qy 58 CTCTTTCTATGTTTCTCTCTGCTGTCCTCCATGCGAGCTCAITTTGGAAGGCAAAAAGGAT 117  
Db 2494 CTCTTTCTGTTGTTCTCTCTTTAGTCTGTTAGTCAITTCAGGTTTGTGCTGCGAGGAG 2435  
Qy 118 AATCAGTTCATCTGGAACAGGTCCTGGGAGGTTACRGGAGACTGTGTCCTCCGGA 177  
Db 2434 TCTTAGTT-GTAGTACACCAAGTTCCTTGGGACTGTTTCAAGAGCTTTTGGTGTGGGT 2376  
Qy 178 GGAGTCCAGAGTCGGGCAGT 197  
Db 2375 TCTGCAGAAAGCTCAGGAGT 2356

Search completed: February 23, 2004, 20:35:49  
Job time : 35.7428 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: February 23, 2004, 20:29:46 ; Search time 12.9894 Seconds  
(without alignments)  
3088.802 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MFPKNLITVCWWSMRKL.....EVKPTAEVCVTAHQHQLHRM 142

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Genesecp1980s:\*
- 2: Genesecp1990s:\*
- 3: Genesecp2000s:\*
- 4: Genesecp2001s:\*
- 5: Genesecp2002s:\*
- 6: Genesecp2003as:\*
- 7: Genesecp2003bs:\*
- 8: Genesecp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	580.5	69.9	1574	7 ADE08053	Ade08053 Novel pro
2	573	69.0	247	4 AAM70792	Aam70792 Human bon
3	573	69.0	247	4 AAM58317	Aam58317 Human bra
4	327.5	39.5	933	7 ADB64774	Adb64774 Human pro
5	297.5	35.8	1624	5 AAE23979	Aae23979 Human Lp2
6	289	34.8	277	4 AAM28984	Aam28984 Peptide #
7	289	34.8	277	4 ABG50341	Abg50341 Human liv
8	289	34.8	617	3 AAB42496	Aab42496 Human ORF
9	253.5	30.5	1549	6 ABR58455	Abr58455 Human NOV
10	252.5	30.4	307	6 ABR58461	Abr58461 Human NOV
11	252.5	30.4	1490	6 ABR58454	Abr58454 Human NOV
12	252.5	30.4	1545	6 ABR58453	Abr58453 Human NOV
13	252.5	30.4	1588	4 AAB20135	Aab20135 Secreted
14	252.5	30.4	1588	4 AAM39295	Aam39295 Human pol
15	252.5	30.4	1588	6 ABR58450	Abr58450 Human NOV
16	252.5	30.4	1588	6 ABR58456	Abr58456 Human NOV
17	252.5	30.4	1588	6 ABR58459	Abr58459 Human NOV
18	252.5	30.4	1588	6 ABR58462	Abr58462 Human NOV
19	128.5	15.5	1057	6 ABP71970	Abp71970 Human agg
20	128.5	15.5	1057	6 ABR40094	Abr40094 Human ADA
21	128.5	15.5	1122	6 ABP71971	Abp71971 Human agg
22	128.5	15.5	1145	6 ABP71974	Abp71974 Human agg
23	127	15.3	622	4 AAB95346	Aab95346 Human pro
24	127	15.3	1363	7 ADD44592	Add44592 Polypepti
25	127	15.3	1505	5 AAU72897	Aau72897 Human met

26	127	15.3	1602	5 ABG30702	Abg30702 Human agg
27	127	15.3	1629	5 ABG30703	Abg30703 Human agg
28	127	15.3	1629	5 AAO14448	Aao14448 Human ADA
29	127	15.3	1907	5 AAU77133	Aau77133 Human pro
30	127	15.3	1916	5 AAE19173	Aae19173 Human pro
31	127	15.3	1934	4 AAB72301	Aab72301 Human ADA
32	127	15.3	1935	5 AAU72896	Aau72896 Human met
33	124	14.9	513	6 ABU54545	Abu54545 Human NOV
34	122	14.7	1882	4 AAB72286	Aab72286 Human ADA
35	121.5	14.6	226	4 AAE10824	Aae10824 Human gen
36	121.5	14.6	465	5 AAE21004	Aae21004 Human pro
37	121.5	14.6	476	5 AAE21001	Aae21001 Human pro
38	121.5	14.6	807	4 ABG14803	Abg14803 Novel hum
39	121.5	14.6	969	5 AAE21003	Aae21003 Human pro
40	121.5	14.6	980	5 AAE21000	Aae21000 Human pro
41	121.5	14.6	1035	5 ABB98128	Abb98128 Human PMM
42	121.5	14.6	1072	6 ABU12107	Abu12107 Human pro
43	121.5	14.6	1094	5 AAU72900	Aau72900 Human met
44	121.5	14.6	1213	5 AAE21002	Aae21002 Human pro
45	121.5	14.6	1224	5 AAE20999	Aae20999 Human pro

ALIGNMENTS

RESULT 1  
ADE08053  
ID ADE08053 standard; protein; 1574 AA.

XX AC ADE08053;  
XX DT 29-JAN-2004 (first entry)  
XX DE Novel protein (useful for identifying genetic disorders) #208.  
XX KW novel gene; novel protein; tissue marker; molecular weight marker;  
XX KW chromosome marker; genetic disorder.  
XX OS Unidentified.  
XX PN WO2003054152-A2.  
XX PD 03-JUL-2003.  
XX PF 10-DEC-2002; 2002WO-US039555.  
XX PR 10-DEC-2001; 2001US-0339739P.  
XX PR 11-DEC-2001; 2001US-0339453P.  
XX PR 14-MAR-2002; 2002US-0365091P.  
XX PR 12-APR-2002; 2002US-0372381P.  
XX PR 12-APR-2002; 2002US-0372615P.  
XX PR 22-APR-2002; 2002US-00128558.  
XX PR 24-APR-2002; 2002US-0376045P.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
XX Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
XX Ma Y, Wang D, Chen R, Xu C, Boyle BJ;  
XX WPI; 2003-569235/53.  
XX N-PSDB; ADE07142.  
XX New polynucleotides, useful for expressing recombinant proteins for  
XX analysis, characterization or therapeutic use, or as markers for tissues  
XX in which the corresponding protein is preferentially expressed.  
XX Claim 20; SEQ ID NO 1119; 1177pp; English.  
XX The invention comprises the amino acid and coding sequences of novel  
XX proteins. The DNA and protein sequences of the invention are useful as:  
XX markers for tissues in which the corresponding protein is preferentially

CC expressed, as molecular weight markers on gels, as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present amino acid sequence represents a protein  
CC of the invention.  
XX  
SQ Sequence 1574 AA;

Query Match 69.9%; Score 580.5; DB 7; Length 1574;  
Best Local Similarity 99.0%; Pred. No. 2.7e-50;  
Matches 98; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 45 KP-GPWGCTGCGGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDL 103  
DB 10 KPHGPGWGCTGCGGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDL 69  
QY 104 FQWEVSDWHHCVLVPYARGEVKPRTAECVTAQHGLQHRM 142  
DB 70 FQWEVSDWHHCVLVPYARGEVKPRTAECVTAQHGLQHRM 108

RESULT 2  
ID AAM70792  
AC AAM70792;  
DT 06-NOV-2001 (first entry)  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 31098.  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000668.

XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human bone marrow.  
XX  
XX Example 4; SEQ ID NO 31098; 6589pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is a  
XX protein encoded by one of the probes of the invention  
XX  
SQ Sequence 247 AA;

Query Match 69.0%; Score 573; DB 4; Length 247;  
Best Local Similarity 100.0%; Pred. No. 2e-50;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDLFQWE 107  
DB 1 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDLFQWE 60  
QY 108 VSDWHHCVLVPYARGEVKPRTAECVTAQHGLQHRM 142  
DB 61 VSDWHHCVLVPYARGEVKPRTAECVTAQHGLQHRM 95

QY 48 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDLFQWE 107  
DB 1 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDLFQWE 60  
QY 108 VSDWHHCVLVPYARGEVKPRTAECVTAQHGLQHRM 142  
DB 61 VSDWHHCVLVPYARGEVKPRTAECVTAQHGLQHRM 95

RESULT 3  
ID AAM58317  
AC AAM58317;  
DT 05-NOV-2001 (first entry)  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 30422.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX  
OS Homo sapiens.  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000667.

XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains.  
XX  
XX Example 4; SEQ ID NO 30422; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, of  
XX epilepsy and cancers. The present sequence is a protein encoded by one of  
XX the probes of the invention  
XX  
SQ Sequence 247 AA;

Query Match 69.0%; Score 573; DB 4; Length 247;  
Best Local Similarity 100.0%; Pred. No. 2e-50;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDLFQWE 107  
DB 1 PWGRTGDCGPGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSFCRVCDWHSDLFQWE 60  
QY 108 VSDWHHCVLVPYARGEVKPRTAECVTAQHGLQHRM 142  
DB 61 VSDWHHCVLVPYARGEVKPRTAECVTAQHGLQHRM 95



QY	22	LLLSLLL-----SPAHL	EKKNOFTWK	GPNGR	CTGD--CGPGGVQSR	AVWCVFVDGHT	75			
DB	34	LLLLLLLP	AGRAAAQGEA	EAP	TLYLWK	TGPGRCMGDECGGGIQT	93			
QY	76	SHLSNCG	BSNPPKRS	CFRCV	DWHS	DLFOWEVS	DWHHCVLVPYAR	GEVKPRTAECVTAQ	135	
DB	94	TLHTNCK	QAEAPN	QNCFKV	CDWHK	ELYDRL	GPWNR	CPV--ISK	LEK	150
QY	136	HGLQHR	141							
DB	151	EGIQVR	156							
RESULT 5										
AAE23979 standard; protein; 1624 AA.										
XX	AAE23979;									
XX	23-SEP-2002	(first entry)								
DT	Human	LP217	secreted	protein.						
DE										
XX	Human;	secreted	protein;	atherosclerosis;	Alzheimer's	disease;	LP217;			
XX	diabetic	retinopathy;	severe	combined	immunodeficiency;	pancreatitis;				
KW	rheumatoid	arthritis;	colorectal	adenoma;	haemolytic	anaemia;	cancer;			
KW	reperfusion	injury;	arteriosclerosis;	wound	healing;	transgenic	animal;			
KW	gene	therapy;	neoplasm;	transgenic;	psoriasis;	ischaemia;	carcinoma;			
KW	chromosome	7p21-p22.								
XX										
XX	Homo	sapiens.								
OS										
XX	Key	Location/Qualifiers								
FX	Peptide	1..15								
FT	/label=	Signal_peptide								
FT	Protein	16..1624								
FT	/note=	"Mature human LP217 secreted protein"								
XX	WO200226801-A2.									
PN										
XX	04-APR-2002.									
PD										
XX	14-SEP-2001;	2001WO-US026026.								
XX										
PF	28-SEP-2000;	2000US-0236088P.								
XX	(ELIL )	LILLY & CO ELI.								
PA										
XX	Su EW,	Wang H;								
PI										
XX	WPI;	2002-471259/50.								
XX	N-PSDB;	AAD38694.								
DR										
XX	Novel	proteins	and	polynucleotides	of	secreted	proteins	useful	for	
PT	treating	various	diseases	e.g.	rheumatoid	arthritis,	cancer,	psoriasis,		
PT	diabetic	retinopathy,	arteriosclerosis,	ischemia	or	reperfusion	injury.			
XX	Claim	8;	Page	117-123;	145pp;	English.				
PS										
XX	The	invention	relates	to	human	secreted	polypeptides	designated	LP095,	
CC	LP191,	LP217,	LP220,	LP221,	LP222,	LP229,	LP237	or	LP238	and
CC	molecules	encoding	such	polypeptides.	Novel	secreted	proteins	of	the	
CC	invention	are	used	for	treating	diseases	such	as	atherosclerosis,	
CC	Alzheimer's	disease,	diabetic	retinopathy,	psoriasis,	pancreatitis,				
CC	arteriosclerosis,	rheumatoid	arthritis,	colorectal	adenomas,	severe				
CC	combined	immunodeficiency,	ischaemia,	carcinomas,	haemolytic	anaemia,				
CC	reperfusion	injury,	neoplasms	and	cancer	especially	liver	cancer.	They	
CC	are	also	used	for	wound	healing.	Polynucleotides	of	the	invention
CC	are	used	to	generate	transgenic	animals	or	knock	out	animals,
CC	are	useful	in	the	development	and	screening	of	therapeutically	useful
CC	reagents	for	use	in	the	treatment	of	diseases	associated	with
CC	polypeptide	associated	activity.	They	are	also	used	in	gene	therapy.
CC	present	sequence	is	human	LP217	secreted	protein.	LP217	gen	is
CC	located									

```
CC on chromosome 7p21-p22
XX Sequence 1624 AA;
SQ Query Match 35.8%; Score 297.5; DB 5; Length 1624;
Best Local Similarity 38.8%; Pred. No. 2.6e-21;
Matches 52; Conservative 22; Mismatches 27; Indels 33; Gaps 5;

QY 9 VTCWVRSMRKLFLLSLLSHAAHLEKGNQFIWKPGPWGRTGD-CGPGGVQSRVAVW 67
DB 22 VPCNI-----FIL-----TGPNRCMGDECGPGGIQTRAVW 52
QY 68 CFHYDGTWTHSLNCGESNRPPKERSCFRVCMDHSDLFOWEVSMDHHCVLVYPYARGEVKPR 127
DB 53 CAHVEGTTTLTNCQAERPNQCNCFKVCMDHKLKELYDRLGPNWQCQPV-ISKSLKRP-110
QY 128 TAEVCVTAOGLQHR 141
DB 111 -LECIKGEGIQVR 123

RESULT 6
AAM28984
ID AAM28984 standard; protein; 277 AA.
XX
AC AAM28984;
XX
XX 17-OCT-2001 (first entry)
XX
DE Peptide #3021 encoded by probe for measuring placental gene expression.
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW Genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488997/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX Claim 27; SEQ ID NO 29253; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs:
XX see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
XX Sequence 277 AA;
XX
XX Query Match 34.8%; Score 289; DB 4; Length 277;
XX Best Local Similarity 48.4%; Pred. No. 2.8e-21;
XX Matches 46; Conservative 20; Mismatches 25; Indels 4; Gaps 3;

QY 48 PWGRCCTGD-CGPGGVQSRVAVWCFHYDGTWTHSLNCGESNRPPKERSCFRVCMDHSDLFOW 106
DB 1 PWGRCMGDECGPGGIQTRAVWCAHVEGTTTLTNCQAERPNQCNCFKVCMDHKLKELYD 60
QY 107 EYSDMHHCVLVYPYARGEVKPRTAECVTAOGLQHR 141
DB 61 RLGPWQCQPV-ISKSLKRP--LECIKGEGIQVR 92

RESULT 7
ABG50341
ID ABG50341 standard; peptide; 277 AA.
XX
AC ABG50341;
XX
XX 25-FEB-2003 (first entry)
XX
XX Human liver peptide, SEQ ID NO 28989.
XX
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX hypercholesterolaemia; coronary heart disease.
XX
XX Homo sapiens.
XX
XX WO200157273-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000664.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488998/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX
XX Claim 27; SEQ ID NO 28989; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 277 AA;
XX
XX Query Match 34.8%; Score 289; DB 4; Length 277;
XX Best Local Similarity 48.4%; Pred. No. 2.8e-21;
XX Matches 46; Conservative 20; Mismatches 25; Indels 4; Gaps 3;

QY 48 PWGRCCTGD-CGPGGVQSRVAVWCFHYDGTWTHSLNCGESNRPPKERSCFRVCMDHSDLFOW 106
DB 1 PWGRCMGDECGPGGIQTRAVWCAHVEGTTTLTNCQAERPNQCNCFKVCMDHKLKELYD 60
```

db 1 PWGRMGDECGPGGIQTRAVWCAHVEGWTTLTHTNCKQAEPRNNQNCFCVCDWHKELYDW 60  
107 EVSDWHHCVLVPARGCVKPRTAECVTAQHGLQHR 141  
61 RLGPWNCQCPV-ISKSLKRP--LECIKGEGIQVR 92

RESULT 8  
AB42496  
ID AAB42496 standard; protein; 617 AA.  
AC AAB42496;  
CX  
CX  
CX 08-FEB-2001 (first entry)  
DT  
DE Human ORFX ORF2260 polypeptide sequence SEQ ID NO:4520.  
DX  
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antianemic; Gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.  
CX  
CX Homo sapiens.  
CX  
CX WO200058473-A2.  
CX  
CX 05-OCT-2000.  
CX  
CX 31-MAR-2000; 2000WO-US008621.  
CX  
CX 31-MAR-1999; 99US-0127607P.  
CX 02-APR-1999; 99US-0127636P.  
CX 05-APR-1999; 99US-0127728P.  
CX 30-MAR-2000; 2000US-00540763.  
CX  
CX (CURA-) CURAGEN CORP.  
CX  
CX Shimkets RA, Leach M;  
CX  
CX WPI; 2000-602362/57.  
CX N-PSDB; AAC76705.  
CX  
CX Novel nucleic acids and peptides derived from open reading frame X,  
CX useful for treating e.g. cancers, proliferative disorders,  
CX neurodegenerative disorders and cardiovascular disease.  
CX  
CX Claim 11; Page 3720-3721; 5507pp; English.  
CX  
CX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CX which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CX sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CX antipsoriatic; antiparkinsonian; neurotropic; neuroprotective; osteopathic;  
CX anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
CX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
CX dermatological; immunosuppressive; antiinflammatory; antibacterial;  
CX antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The  
CX sequences can be used for determining the presence of or predisposition  
CX to, or preventing or treating pathological conditions associated with an  
CX ORFX-associated disorder. The nucleic acids can be used to express ORFX  
CX proteins in gene therapy vectors. The proteins and nucleic acids may be  
CX used to treat cancers, proliferative disorders, neurodegenerative  
CX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
CX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester

CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive  
XX  
SQ Sequence 617 AA;  
Query Match 34.8%; Score 289; DB 3; Length 617;  
Best Local Similarity 48.4%; Pred. No. 6,7e-21;  
Matches 46; Conservative 20; Mismatches 25; Indels 4; Gaps 3;  
QY 48 PWGRCTGD-CGPGVQSRVWCFHVDGWTSHLNGESNPPKERSCFRVCDSHDLFW 106  
DB 1 PWGRMGDECGPGGIQTRAVWCAHVEGWTTLTHTNCKQAEPRNNQNCFCVCDWHKELYDW 60  
QY 107 EVSDWHHCVLVPARGCVKPRTAECVTAQHGLQHR 141  
DB 61 RLGPWNCQCPV-ISKSLKRP--LECIKGEGIQVR 92

RESULT 9  
ABR58455  
ID ABR58455 standard; protein; 1549 AA.  
AC ABR58455;  
CX  
CX 07-JUL-2003 (first entry)  
DT  
DE Human NOVA7f.  
DX  
KW Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytostatic; neurotropic; neuroprotective; dyslipidaemia;  
KW antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;  
KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
KW neurodegenerative disorder; Alzheimer's disease; immune disorder;  
KW haematopoietic disorder.  
CX  
CX Homo sapiens.  
CX  
CX WO2003029423-A2.  
CX 10-APR-2003.  
CX  
CX 02-OCT-2002; 2002WO-US031358.  
CX  
CX 02-OCT-2001; 2001US-0326483P.  
CX 05-OCT-2001; 2001US-0327342P.  
CX 09-OCT-2001; 2001US-0327917P.  
CX 09-OCT-2001; 2001US-0328029P.  
CX 09-OCT-2001; 2001US-0328044P.  
CX 09-OCT-2001; 2001US-0328056P.  
CX 12-OCT-2001; 2001US-0328849P.  
CX 15-OCT-2001; 2001US-0329414P.  
CX 17-OCT-2001; 2001US-0330142P.  
CX 22-OCT-2001; 2001US-0341058P.  
CX 24-OCT-2001; 2001US-0343629P.  
CX 29-OCT-2001; 2001US-0349575P.  
CX 01-NOV-2001; 2001US-0346357P.  
CX 12-APR-2002; 2002US-0371972P.  
CX 12-APR-2002; 2002US-0371980P.  
CX 17-APR-2002; 2002US-0373261P.  
CX 19-APR-2002; 2002US-0373805P.  
CX 23-APR-2002; 2002US-0374738P.  
CX 16-MAY-2002; 2002US-0381101P.  
CX 17-MAY-2002; 2002US-0381635P.  
CX 21-MAY-2002; 2002US-0383830P.  
CX 01-OCT-2002; 2002US-00262839.  
CX (CURA-) CURAGEN CORP.  
CX  
CX Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;

PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
 PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
 PI Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;  
 PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
 XX WPI: 2003-381625/36.  
 DR N-PSDB; ACC72167.  
 XX  
 XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
 PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
 PT dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 PS Claim 1; Page 260; 487pp; English.  
 XX  
 XX The present invention relates to novel human NOV proteins and their  
 CC coding sequences (ACC72075-ACC72181 and ABR58461). The NOV  
 CC proteins are useful in manufacturing a medicament for treating a syndrome  
 CC associated with a human disease. The NOV proteins and coding sequences  
 CC may be used to diagnose, treat or prevent metabolic disorders such as  
 CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
 CC disorders such as Alzheimer's disease or Parkinson's disease, immune  
 CC disorders, haematopoietic disorders and various dyslipidaemias  
 XX  
 SQ Sequence 1549 AA;  
 Query Match 30.5%; Score 253.5; DB 6; Length 1549;  
 Best Local Similarity 43.5%; Pred. No. 8e-17; Mismatches 28; Indels 3; Gaps 2;  
 Matches 40; Conservative 21; Mismatches 28; Indels 3; Gaps 2;  
 QY 50 GRCTGCGGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKERSCFRVCDSDFOWEVS 109  
 DB 1 GTMGDECGGGIQTAVWCAHVEGWTTLTNCQAERNQNCFKVCDWHKELYDRLG 60  
 QY 110 DWHHCULPYARGEVKPTACVTAHQGLQHR 141  
 DB 61 PWNQCQPV-ISKLEKP--LECIRGEEGIQVR 89  
 RESULT 10  
 ABR58461  
 ID ABR58461 standard; protein; 307 AA.  
 AC ABR58461;  
 XX  
 XX 07-JUL-2003 (first entry)  
 DT Human NOV47L.  
 DE  
 XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;  
 KW antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;  
 KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; immune disorder;  
 KW haematopoietic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO2003029423-A2.  
 XX  
 PD 10-APR-2003.  
 XX  
 FF 02-OCT-2001; 2002WO-US031358.  
 XX  
 XX 02-OCT-2001; 2001US-0326483P.  
 PR 05-OCT-2001; 2001US-0327342P.  
 PR 09-OCT-2001; 2001US-0327917P.  
 PR 09-OCT-2001; 2001US-0328029P.  
 PR 09-OCT-2001; 2001US-0328044P.  
 PR 09-OCT-2001; 2001US-0328056P.  
 PR 12-OCT-2001; 2001US-0328849P.  
 PR 15-OCT-2001; 2001US-0329414P.  
 PR 17-OCT-2001; 2001US-0330142P.

PR 22-OCT-2001; 2001US-0341058P.  
 PR 24-OCT-2001; 2001US-0339266P.  
 PR 24-OCT-2001; 2001US-0343629P.  
 PR 29-OCT-2001; 2001US-0349575P.  
 PR 01-NOV-2001; 2001US-0346357P.  
 PR 12-APR-2002; 2002US-0371972P.  
 PR 12-APR-2002; 2002US-0371980P.  
 PR 17-APR-2002; 2002US-0373261P.  
 PR 19-APR-2002; 2002US-0373805P.  
 PR 23-APR-2002; 2002US-0374738P.  
 PR 16-MAY-2002; 2002US-0381101P.  
 PR 17-MAY-2002; 2002US-0381635P.  
 PR 29-MAY-2002; 2002US-0383830P.  
 PR 01-OCT-2002; 2002US-00262839.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Alsbrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
 PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
 PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
 PI Rothenberg ME, Shimkets RA, Smithson G, Spytek KA, Taupier RJ;  
 PI Vernet CAM, Voss EZ, Zerhusen BD, Zhong M;  
 XX WPI: 2003-381625/36.  
 DR N-PSDB; ACC72173.  
 XX  
 XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
 PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
 PT dyslipidemia, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics.  
 XX  
 PS Claim 1; Page 266; 487pp; English.  
 XX  
 XX The present invention relates to novel human NOV proteins and their  
 CC coding sequences (ACC72075-ACC72181 and ABR58469). The NOV  
 CC proteins are useful in manufacturing a medicament for treating a syndrome  
 CC associated with a human disease. The NOV proteins and coding sequences  
 CC may be used to diagnose, treat or prevent metabolic disorders such as  
 CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
 CC disorders such as Alzheimer's disease or Parkinson's disease, immune  
 CC disorders, haematopoietic disorders and various dyslipidaemias  
 XX  
 XX Sequence 307 AA;  
 Query Match 30.4%; Score 252.5; DB 6; Length 307;  
 Best Local Similarity 44.8%; Pred. No. 1.7e-17; Mismatches 24; Indels 3; Gaps 2;  
 Matches 39; Conservative 21; Mismatches 24; Indels 3; Gaps 2;  
 QY 55 DCGPGVQSVRAVWCFHVDGWTSHLSNCGESNRPKERSCFRVCDSDFOWEVSDFHHC 114  
 DB 4 ECGPGGIQTAVWCAHVEGWTTLTNCQAERNQNCFKVCDWHKELYDRLGPNQC 63  
 QY 115 VLVYPARGEVKPTACVTAHQGLQHR 141  
 DB 64 QPV-ISKLEKP--LECIRGEEGIQVR 87  
 RESULT 11  
 ABR58454  
 ID ABR58454 standard; protein; 1490 AA.  
 XX  
 AC ABR58454;  
 XX  
 XX 07-JUL-2003 (first entry)  
 DT Human NOV47e.  
 DE  
 XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
 KW immunomodulator; cytostatic; nootropic; neuroprotective; dyslipidaemia;  
 KW antiparkinsonian; antilipaeamic; gene therapy; metabolic disorder;  
 KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; immune disorder;  
 KW haematopoietic disorder.



The present sequence is that of ovel secreted protein SSCP1, which is predicted to localise in the plasma membrane. The protein shows homology to human KIAA0960 protein. RNA species with homology to the SSCP1 cDNA were detected in endometrial cells, pancreas, adipose, adrenal gland, thyroid, mammary gland, myometrium, uterus, placenta, prostate testis, and in neoplastic cells derived from ovarian carcinoma, breast carcinoma and prostate carcinoma (bone metastases) and melanoma. SSCP1 was also highly expressed in a small cell lung cancer, a large cell lung cancer and a

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AA38642-AA42213) with nootropic, immunosuppressant and cytoskeletal activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Actinin/inhibin activity, chemoractic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders. Note: The sequence data for this patent did not form part of the printed specification

XX Sequence 1588 AA;

Query Match 30.4%; Score 252.5; DB 4; Length 1588;  
Best Local Similarity 44.8%; Pred. No. 1e-16;  
Matches 39; Conservative 21; Mismatches 24; Indels 3; Gaps 2;

QY 55 DCGPGGVQSRVAVWCFHVDGWTSHLSNCGESNRPKERSCFRVCDSHSDLFOWEVSVDWHHC 114  
Db 4 ECGPGGIQTRAVWCAHVEGWTTLHTNCKQAEPRNNQCNCFKVCDSHSDLFOWEVSVDWHHC 63

QY 115 VLVPYARGEVKPRTAEVCVTAQHGLQHR 141  
Db 64 QPV-ISKSLERP--LECIKGEEGIQVR 87

## RESULT 15

ID ABR58450 standard; protein; 1588 AA.

XX ABR58450;

XX AC

XX 07-JUL-2003 (first entry)

XX Human NOV47a.

XX Human; NOV; antidiabetic; anorectic; antibacterial; virucide;  
KW immunomodulator; cytoskeletal; nootropic; neuroprotective; dyslipidaemia;  
KW antiparkinsonian; antilipidemic; gene therapy; metabolic disorder;  
KW diabetes; obesity; infection; cachexia; cancer; Parkinson's disease;  
KW neurodegenerative disorder; Alzheimer's disease; immune disorder;  
KW haematopoietic disorder.

XX Homo sapiens.

XX WO2003029423-A2.

XX 10-APR-2003.

XX 02-OCT-2002; 2002WO-US031358.

XX 02-OCT-2001; 2001US-0326483P.

XX 05-OCT-2001; 2001US-0327342P.

XX 09-OCT-2001; 2001US-0327917P.

XX 09-OCT-2001; 2001US-0328029P.

XX 09-OCT-2001; 2001US-0328044P.

XX 09-OCT-2001; 2001US-0328056P.

XX 12-OCT-2001; 2001US-0328499P.

XX 15-OCT-2001; 2001US-0329414P.

XX 17-OCT-2001; 2001US-0330142P.

XX 22-OCT-2001; 2001US-0341058P.

XX 24-OCT-2001; 2001US-0339266P.

XX 24-OCT-2001; 2001US-0343629P.

XX 29-OCT-2001; 2001US-0349575P.

XX 01-NOV-2001; 2001US-0346357P.

XX 12-APR-2002; 2002US-0371972P.

XX 12-APR-2002; 2002US-0371980P.

XX 17-APR-2002; 2002US-0373261P.

PR 19-APR-2002; 2002US-0373805P.  
PR 23-APR-2002; 2002US-0374738P.  
PR 16-MAY-2002; 2002US-0381101P.  
PR 17-MAY-2002; 2002US-0381635P.  
PR 29-MAY-2002; 2002US-0383830P.  
PR 01-OCT-2002; 2002US-00262839.

XX (CURA-) CURAGEN CORP.

PI Alsebrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
PI Edinger SR, Ellerman K, Gerlach VL, Gorman L, Guo X, Ji W;  
PI Kekuda R, Leach MD, Li L, Miller CE, Patturajan M, Rieger DK;  
PI Rothenberg ME, Shmukler RA, Smithson G, Spytex KA, Taupier RJ;  
PI Vernet CM, Voss EZ, Zerhusen BD, Zhong M;  
XX WPI; 2003-381625/36.  
DR N-PSDB; ACC72162.

XX NOVX polypeptides and nucleic acids useful for diagnosing, preventing or  
PT treating NOVX-associated disorders, e.g. diabetes, obesity, cancer or  
PT dyslipidemia, and in chromosome mapping, tissue typing or  
PT pharmacogenomics.

XX Claim 1; Page 254-255; 487pp; English.

XX The present invention relates to novel human NOV proteins and their  
CC coding sequences (ACC72075-ACC72181 and ABR58363-ABR58469). The NOV  
CC proteins are useful in manufacturing a medicament for treating a syndrome  
CC associated with a human disease. The NOV proteins and coding sequences  
CC may be used to diagnose, treat or prevent metabolic disorders such as  
CC diabetes or obesity, infections, cachexia, cancer, neurodegenerative  
CC disorders such as Alzheimer's disease or Parkinson's disease, immune  
CC disorders, haematopoietic disorders and various dyslipidaemias

XX Sequence 1588 AA;

Query Match 30.4%; Score 252.5; DB 6; Length 1588;  
Best Local Similarity 44.8%; Pred. No. 1e-16;

Matches 39; Conservative 21; Mismatches 24; Indels 3; Gaps 2;

QY 55 DCGPGGVQSRVAVWCFHVDGWTSHLSNCGESNRPKERSCFRVCDSHSDLFOWEVSVDWHHC 114  
Db 4 ECGPGGIQTRAVWCAHVEGWTTLHTNCKQAEPRNNQCNCFKVCDSHSDLFOWEVSVDWHHC 63

QY 115 VLVPYARGEVKPRTAEVCVTAQHGLQHR 141

Db 64 QPV-ISKSLERP--LECIKGEEGIQVR 87

Search completed: February 24, 2004, 01:02:09

Job time : 13.9894 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 23, 2004, 17:26:11 ; Search time 140.513 Seconds  
(without alignments)  
10566.289 Million cell updates/sec

Title: US-10-022-710-5\_COPY\_1\_424

Perfect score: 424

Sequence: 1 atgttccaaagacacact.....gcattgacgacgacacgga 424

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	285	67.2	744	9	US-09-864-761-27172
2	264.4	62.4	477	9	US-09-864-761-10551
3	78.6	18.5	3053	15	US-10-104-047-958
4	76	17.9	6378	14	US-10-037-270-332
5	76	17.9	6378	15	US-10-117-722-332
6	75.4	17.8	1956	9	US-09-864-761-13883
7	75.4	17.8	6373	13	US-10-000-512-1
8	75.4	17.8	6373	15	US-10-074-566-1
9	75.4	17.8	6373	15	US-10-074-566-58
10	73.8	17.4	833	9	US-09-864-761-30452
11	73.8	17.4	4810	15	US-10-074-566-40
12	35.8	8.4	1629	9	US-09-732-224-6
13	34.2	8.1	652	15	US-10-027-632-232714
14	34.2	8.1	82121	15	US-10-065-111-136
15	34.2	8.1	358246	15	US-10-292-798-1095

C 15	33.6	7.9	2939	15	US-10-080-334-69	Sequence 69, Appl
C 17	32.4	7.6	794	15	US-10-027-632-156348	Sequence 156348,
18	32.4	7.6	1102	12	US-10-388-703-28	Sequence 28, Appl
19	32.4	7.6	1103	12	US-10-388-703-26	Sequence 26, Appl
20	32.4	7.6	1103	12	US-10-388-703-27	Sequence 27, Appl
21	32.4	7.6	1103	12	US-10-388-703-30	Sequence 30, Appl
22	32.4	7.6	2598	9	US-09-764-887-432	Sequence 432, App
C 23	32.4	7.6	2598	14	US-10-073-961-432	Sequence 432, App
24	32	7.5	440	14	US-10-184-644-386	Sequence 386, App
25	32	7.5	440	14	US-10-184-644-386	Sequence 386, App
26	32	7.5	440	14	US-10-063-685-134	Sequence 134, App
C 27	32	7.5	1957	14	US-10-198-846-9708	Sequence 9708, Ap
C 28	31.8	7.5	588	15	US-10-027-632-105925	Sequence 105925,
C 29	31.8	7.5	588	15	US-10-027-632-131222	Sequence 131222,
C 30	31.8	7.5	588	15	US-10-027-632-131223	Sequence 131223,
C 31	31.8	7.5	588	15	US-10-027-632-131224	Sequence 131224,
C 32	31.8	7.5	722	15	US-10-027-632-19326	Sequence 19326, A
33	31.8	7.5	32127	15	US-10-242-355-1019	Sequence 1019, Ap
C 34	31.4	7.4	2674	9	US-09-917-800A-1500	Sequence 1500, Ap
C 35	31.4	7.4	2674	15	US-10-388-934-584	Sequence 584, App
C 36	31.4	7.4	8113	14	US-10-072-900-4	Sequence 4, Appli
C 37	31.4	7.4	8113	14	US-10-072-900-3	Sequence 3, Appli
C 38	31.4	7.4	8875	14	US-10-072-900-2	Sequence 2, Appli
C 39	31.4	7.4	9112	14	US-10-072-900-1	Sequence 1, Appli
40	31.2	7.4	521	15	US-10-027-632-264412	Sequence 264412,
41	31.2	7.4	1128	14	US-10-156-761-6651	Sequence 6651, Ap
C 42	31.2	7.4	203654	9	US-09-820-905-3	Sequence 3, Appli
C 43	31.2	7.4	418550	15	US-10-292-798-1463	Sequence 1463, Ap
44	31.2	7.4	9025608	14	US-10-156-761-1	Sequence 1, Appli
45	31	7.3	2866	10	US-09-866-050A-598	Sequence 598, App

#### ALIGNMENTS

#### RESULT 1

US-09-864-761-27172  
; Sequence 27172, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30



; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 27172  
; LENGTH: 744  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC011231.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
; OTHER INFORMATION: EST\_HUMAN HIT: BE295719.1, EVALU6 6.00e-53  
; OTHER INFORMATION: NT HIT: AF133643.1, EVALU6 1.90e-00  
; OTHER INFORMATION: SW-SSPROT HIT: P35446, EVALU6 6.00e-10  
US-09-864-761-27172

Query Match 67.2%; Score 285; DB 9; Length 744;  
Best Local Similarity 100.0%; Pred. No. 2.4e-88;  
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 140 GTCGCGGGAAGGTGTACAGAGACTGTGTCGCGAGAGTCCAGAGTCCAGAGTGGGCGAGTGT 199  
DB 1 GTCGCGGGAAGGTGTACAGAGACTGTGTCGCGAGAGTCCAGAGTGGGCGAGTGT 60  
QY 200 GGTGTTTTCATGTCGCGGTGCAGAGTCACTGTCTAATCTGCTGAGAGCAACAGC 259  
DB 61 GGTGTTTTCATGTTGACGGGTGCAGAGTCACTGTCTAATCTGCTGAGAGCAACAGC 120  
QY 260 CTCGAAGGAAGAAGTGTGTTCCGAGTTGTGACTGGCAGAGTCACTCTTTCAAGTGGG 319  
DB 121 CTCGAAGGAAGAAGTGTGTTCCGAGTTGTGACTGGCAGAGTCACTCTTTCAAGTGGG 180  
QY 320 AGTTTTCAGTGGACCACTGTGCTGCTTACCTTACCTGCGGTGAAGTCAAGCTTC 379  
DB 181 AGTTTTCAGTGGACCACTGTGCTGCTTACCTTACCTGCGGTGAAGTCAAGCTTC 240  
QY 380 GGACTGCAGAGTGTGACGGCTCAGCATGGACTGCAGCACCGGA 424  
DB 241 GGACTGCAGAGTGTGACGGCTCAGCATGGACTGCAGCACCGGA 285

RESULT 2  
US-09-864-761-10551  
; Sequence 10551, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Reomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 10551  
; LENGTH: 477  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC011231.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.89  
US-09-864-761-10551

Query Match 62.4%; Score 264.4; DB 9; Length 477;  
Best Local Similarity 85.3%; Pred. No. 2.9e-81;  
Matches 295; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
QY 53 GGAAGCTCTTTCTATTGCTTCTCTCTGCTGCCATCGAGCTCATTTGGAGGCAAAA 112  
DB 132 GGAAGTGTGTTAATTGACTTCTACTTACCTGTGAATCTCTGAATCTCTGAGTAA 191  
QY 113 AGGATAATCAGTTTCATCTGGAACACAGGTCGCTGGGAAAGGTGTACAGGAGACTGTGGTC 172  
DB 192 TTAACATCTTGTGTTTCTGCTGTAGTCCGTGGGAGGTGTACAGGAGACTGTGGTC 251  
QY 173 CCGGAGAGTCCAGAGTCCGGCAGTGTGTTTTCATGTTGACGGGTGACCAAGTCACC 232  
DB 252 CCGGAGAGTCCAGAGTCCGGCAGTGTGTTTTCATGTTGACGGGTGACCAAGTCACC 311  
QY 233 TGTCTAACTGCTGTGAGAGCAACAGAGCTCCAAAGGAAGAAGTGTGTTCCGAGTTTGTG 292  
DB 312 TGTCTAACTGCTGTGAGAGCAACAGAGCTCCAAAGGAAGAAGTGTGTTCCGAGTTTGTG 371  
QY 293 ACTGGCAGTGCAGCTCTTTTCAGTGGAGGTTTCTGACTGGCAGGCTGCTGCTGCTGCTC 352  
DB 372 ACTGGCAGTGCAGCTCTTTTCAGTGGAGGTTTCTGACTGGCAGGCTGCTGCTGCTGCTC 431  
QY 353 CTTACGCTCCGGTGAAGTCAAGCTCGGACTCGAGAGTGTGTGAC 398  
DB 432 CTTACGCTCCGGTGAAGTCAAGCTCGGACTCGAGAGTGTGTGAC 477

RESULT 3  
US-10-104-047-958  
; Sequence 958, Application US/10104047

Publication No. US20030236392A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20030236392A1el full length cdna  
FILE REFERENCE: HI-A0105  
CURRENT APPLICATION NUMBER: US/10/104,047  
CURRENT FILING DATE: 2002-03-25  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 4096  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 958  
LENGTH: 3053  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-104-047-958

Query Match 18.5%; Score 78.6; DB 15; Length 3053;  
Best Local Similarity 58.4%; Pred. No. 2.7e-16;  
Matches 157; Conservative 0; Mismatches 109; Indels 3; Gaps 1;  
2Y 90 TGCAGCTCATTTGGAGGCAAAAGGATATATCAGTTTCATCTGGAACCAAGGTCCTGGGG 149  
Db 393 TGGGGCGCAGGGCGAGGGCGGCCACCTCTATCTGTGAAGACTGGTCCATGGGG 452  
2Y 150 AAGGTGTACAGGA--GACTGTGTCCCGAGGAGTCCAGAGTCCGAGCTGGGCTGGTGT 206  
Db 453 CCAGTGTATGGAGATGATGTGGTCCCGAGGATCCAAACGAGGGCTGTGGTGTGC 512  
2Y 207 TCATGTTCAGGGTGGACAAAGTCACTGTCTACTGTGTGAGAGCAACAGGGCTCCAAA 266  
Db 513 TCATGTGGAGGATGACTACACTGCATCTACTGTGAAGCAGCGGAGAGCCCAATAA 572  
2Y 267 GGAAGAGATGTTTCCGAGTTTGTGACTGGCACAGTACCTTTTCAGTGGAGGTTTC 326  
Db 573 CCAGCAGATGTTTCAAAGTTTGCATTGGCAAAAGAGTTGTACGACTGGAGACTGGG 632  
2Y 327 TGACTGGCACCACCTGTGTGTGTCTTCTT 355  
Db 633 ACCTTGGATCGGTGTCAGGCCCGTGATT 661

RESULT 4  
US-10-037-270-332  
Sequence 332, Application US/10037270  
Publication No. US20030104529A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aigong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunqing  
APPLICANT: Wang, Duanrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: Tillinghast, John  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/10/037,270  
CURRENT FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21

NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 332  
LENGTH: 6378  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (271)..(5037)  
NAME/KEY: misc feature  
LOCATION: (1)..(6378)  
OTHER INFORMATION: n = a,t,c or g  
US-10-037-270-332

Query Match 17.9%; Score 76; DB 14; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 3.1e-15;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;  
QY 115 GATAATCAGTTCATCTCGAATACCAAGTCCGTCGGGAGAGGTGTACAGGA---GACTGTGGT 171  
Db 229 GGTAATTGGGAAAGCTTCTTGAGAAAGTCCATGGGGCCGATGTATGGAGATGAATGTGGT 288  
QY 172 CCGGAGGAGTCCAGAGTCGGGCGAGTGTGGTGTGTTTCATGTTGACGGTGGACAAAGTCAC 231  
Db 289 CCGGAGGCATCCAAACGAGGGCTGTGTGTGTCTCATGTGGAGGATGCACTACACTG 348  
QY 232 CTGTCTAACTGTGTGTGAGAGCAACAGGCTCCAAAGAAAGAGTTGTTCCGAGTTGT 291  
Db 349 CATACTAATCTGAAGCAGCGCCGAGAGACCCCAATAACACAGCAGAGATTGTTTCAAAGTTGC 408  
QY 292 GACTGGCACAGTGCACCTTTTCAGTGGGAGGTTTCTCACTGGCACCACCTGTGTGCTTGT 351  
Db 409 GATTGGCCAAAGAGTTGTACGACTGGAGCTGGGACCTTGGAAATCATGTGTCACCCGTG 468  
QY 352 CCTT 355  
Db 469 ATTT 472

RESULT 5  
US-10-117-722-332  
Sequence 332, Application US/10117722  
Publication No. US20030219744A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and  
FILE REFERENCE: 784CIP2BCIP  
CURRENT APPLICATION NUMBER: US/10/117,722  
CURRENT FILING DATE: 2002-04-04  
PRIOR APPLICATION NUMBER: 09/620,312  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 332  
LENGTH: 6378  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (271)..(5037)  
NAME/KEY: misc feature  
LOCATION: (1)..(6378)  
OTHER INFORMATION: n = a,t,c or g

US-10-117-722-332

Query Match 17.9%; Score 76; DB 15; Length 6378;  
Best Local Similarity 59.8%; Pred. No. 3.1e-15;  
Matches 146; Conservative 0; Mismatches 95; Indels 3; Gaps 1;

QY 115 GATAATCAGTTCATCTGAAACCCAGGTCCTGGGGAAGGTGTACAGGA---GACTGTGGT 171  
DB 229 GGTAAATTGGAAAGCTTCTGAGAAAGTCCATGGGCGCATGTATGGAGATGAATGTGGT 288  
QY 172 CCCGAGGAGTCCAGAGTCGGGCACTGTGGTGTTCATGTTCAGCGGTGACAAAGTCA 231  
DB 289 CCCGAGGAGTCCAAACAGGCGCTGTGGTGTCTCATGTGGAGGATGACTACACTG 348  
QY 232 CTGTCTAACTGTGGTGAGAGCAACAGGCTCCAAAGGAAGAGTGTTCCTCCAGTTGT 291  
DB 349 CATACTTAAGTAAACAGGCGGAGAGCCCAATACAGCAGAGATTTTCAAGATTGC 408  
QY 292 GACTGGCAGTGCCTTTTCAGTGGGAGGTTTCTGACTGGCACCAGTGTGTCTTGT 351  
DB 409 GATTGGCAAAAGAGTTGTACGACTGGAGACTGGGACCTTGAATCAGTGTACGCCGCTG 468  
QY 352 CCTT 355  
DB 469 ATTT 472

## RESULT 6

US-09-864-761-13883  
; Sequence 13883, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 13883  
; LENGTH: 1956  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC004614.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9

US-09-864-761-13883

Query Match 17.8%; Score 75.4; DB 9; Length 1956;  
Best Local Similarity 62.0%; Pred. No. 2.9e-15;  
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
QY 138 AGTCCGTGGGGAAGGTGTACAGGA---GACTGTGTGTCCTCCGAGGAGTCCAGAGTCGGGC 194  
DB 189 AGTCCATGGGCCGATGTATGGAGATGAATGTGTCCTCCGAGGATCCAAACGAGGC 248  
QY 195 AGTGTGTGTTTCATGTTCAGCGGTGGACAAAGTCACTGTCTAAGTGTGTGAGAGCA 254  
DB 249 TGTGTGTGTCATGTGAGGGATGGACTACCTGCATACCTAACTGTAAAGCAGGCCGA 308  
QY 255 CAGGCTCCAAAGGAAGTGTTCCTGAGTTGTGAGTGGCAGAGTCACTCTTTCA 314  
DB 309 GAGACCAATACCAGCAGAAATTTTCAAGTTTTCGATTGCGACAAAGAGTTGTACGA 368  
QY 315 GTGGAGGTTTCTGACTGGCACCACCTGTGTGTGTTCCTT 355  
DB 369 CTGGAGACTGGGACCTTGAATCAGTGTGAGCCCGTGATT 409

## RESULT 7

US-10-000-512-1  
; Sequence 1, Application US/10000512  
; Publication No. US2002016499A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimketa, Richard A  
; APPLICANT: Fernandes, Elma  
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME  
; FILE REFERENCE: 15966-556  
; CURRENT APPLICATION NUMBER: US/10/000,512  
; CURRENT FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 6373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (235)..(4998)  
; NAME/KEY: misc feature  
; LOCATION: (6349)  
; OTHER INFORMATION: Wherein n is a or t or g or c.  
US-10-000-512-1

Query Match 17.8%; Score 75.4; DB 13; Length 6373;  
Best Local Similarity 62.0%; Pred. No. 4.9e-15;  
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
QY 138 AGTCCGTGGGGAAGGTGTACAGGA---GACTGTGTGTCCTCCGAGGAGTCCAGAGTCGGGC 194  
DB 216 AGTCCATGGGCCGATGTATGGAGATGAATGTGTCCTCCGAGGATCCAAACGAGGC 275

QY 195 AGTGTGTTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGGTGAGAGCAA 254  
Db 276 TGTGTGTGTGCTCATGTGGAGGATGGATCACTGTCATCTAACTGTGAAGCAGGCCGA 335  
QY 255 CAGGCTCCAAAGAAAGAGTTGTTTCCGAGTTTGTGACTGGCAGTGAACCTCTTTCA 314  
Db 336 GAGACCAATAACACGAGAGATTGTTTCAAAGTTTGGATTGGCAAAAGAGTTGTACGA 395  
QY 315 GTGGAGGTTTCTGACTGGCAGCAGTGTGCTGTTCCTT 355  
Db 396 CTGGAGACTGGACCTTGGATCATGTGTACGCCGTGATT 436

RESULT 8  
US-10-074-566-1  
; Sequence 1, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; PRIOR FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 6373  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (6349)  
; OTHER INFORMATION: Wherein N is A, or T, or C, or G.

Query Match 17.8%; Score 75.4; DB 15; Length 6373;  
Best Local Similarity 62.0%; Pred. No. 4.9e-15;  
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
QY 138 AGTCCCTGGGAAAGTGTACAGCA---GACTGTGTGTCCTGGAGAGTCCAGAGTCGGGC 194  
Db 216 AGTCCATGGGGCCGATGTATGGAGATGATGTGTCCTGGAGGCATCCAAACGAGGC 275  
QY 195 AGTGTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGTGAGAGCAA 254  
Db 276 TGTGTGTGCTCATGTGGAGATGATGTGTCCTGGAGGCATCCAAACGAGGC 335  
QY 195 AGTGTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGTGAGAGCAA 254  
Db 276 TGTGTGTGCTCATGTGGAGATGATGTGTCCTGGAGGCATCCAAACGAGGC 335

QY 255 CAGGCTCCAAAGAAAGTGTGTTTCCGAGTTTGTGACTGGCAGTGAACCTCTTTCA 314  
Db 336 GAGACCAATAACACGAGAGATTGTTTCAAAGTTTGGATTGGCAAAAGAGTTGTACGA 395  
QY 315 GTGGAGGTTTCTGACTGGCAGCAGTGTGCTGTTCCTT 355  
Db 396 CTGGAGACTGGACCTTGGATCATGTGTACGCCGTGATT 436

RESULT 9  
US-10-074-566-58  
; Sequence 58, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; PRIOR FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 58  
; LENGTH: 6373  
; TYPE: DNA  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (6349)  
; OTHER INFORMATION: Wherein N is A, or T, or C, or G.

Query Match 17.8%; Score 75.4; DB 15; Length 6373;  
Best Local Similarity 62.0%; Pred. No. 4.9e-15;  
Matches 137; Conservative 0; Mismatches 81; Indels 3; Gaps 1;  
QY 138 AGTCCCTGGGAAAGTGTACAGCA---GACTGTGTGTCCTGGAGAGTCCAGAGTCGGGC 194  
Db 216 AGTCCATGGGGCCGATGTATGGAGATGATGTGTCCTGGAGGCATCCAAACGAGGC 275  
QY 195 AGTGTGTTTTCATGTTGACGGTGGCAAGTCACTGTCTAACTGTGTGAGAGCAA 254  
Db 276 TGTGTGTGCTCATGTGGAGATGATGTGTCCTGGAGGCATCCAAACGAGGC 335  
QY 255 CAGGCTCCAAAGAAAGTGTGTTTCCGAGTTTGTGACTGGCAGTGAACCTCTTTCA 314  
Db 336 GAGACCAATAACACGAGAGATTGTTTCAAAGTTTGGATTGGCAAAAGAGTTGTACGA 395

QY 315 GTGGAGCTTTCTGACTGGCACCAGTGTGCTGTTCCTT 355  
DB 396 CTGGAGACTGGGACCTTGGAAATCAGTGTGACCCCGTGATT 436

RESULT 10  
US-09-864-761-30452  
; Sequence 30452, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 30452  
; LENGTH: 833  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC004614.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9  
; OTHER INFORMATION: SWISSPROT HIT: P35446, EVALUE 3.00e-12  
; OTHER INFORMATION: NT HIT: A3030073.1, EVALUE 1.10e-01  
; OTHER INFORMATION: EST HUMAN HIT: AW612526.1, EVALUE 0.00e+00  
US-09-864-761-30452  
Query Match 17.4%; Score 73.8; DB 9; Length 833;  
Best Local Similarity 60.9%; Pred. No. 7.2e-15;  
Matches 120; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 159 AGGAGACTGTGTCCCGGAGGAGTCCAGAGTCCGCGAGTGTGTTCATGTTGACGG 218  
DB 23 AGATGAATGTGTCCCGGAGGCATCCAAACGAGGGCTGTGTGCTCATGTGGAGGG 82

Matches 120; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 159 AGGAGACTGTGTCCCGGAGGAGTCCAGAGTCCGCGAGTGTGTTCATGTTGACGG 218  
DB 23 AGATGAATGTGTCCCGGAGGCATCCAAACGAGGGCTGTGTGCTCATGTGGAGGG 82  
QY 219 GTGGACAAGTCACTGTCTAACTGTGTGAGAGCAACAGGCCTCCAAAGGAAGAAGTTG 278  
DB 83 ATGGACTACACTGCATACACTGTGTAAGCAGCCCGAGAGACCAATAACCCAGCAGATTG 142  
QY 279 TTTCGAGTTTGTGACTGGCAGAGTACCTCTTTAGTGGGAGGTTTCTGACTGGCACCA 338  
DB 143 TTTCAAAAGTTTGGATTGGCACAAGAGTTGTACGACTGGGACCTTGAATCA 202  
QY 339 CTGTGTGCTTGTTCCTT 355  
DB 203 GTGTGAGCCCGTGATT 219

RESULT 11  
US-10-074-566-40  
; Sequence 40, Application US/10074566  
; Publication No. US20030207348A1  
; GENERAL INFORMATION:  
; APPLICANT: Shinkets, Richard A.  
; APPLICANT: Fernandes, Elma R.  
; APPLICANT: Li, Li  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gusev, Vladimir Y.  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shenoy, Suresh G.  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same  
; FILE REFERENCE: 15966-556 CIP1  
; CURRENT APPLICATION NUMBER: US/10/074,566  
; CURRENT FILING DATE: 2002-02-13  
; PRIOR APPLICATION NUMBER: 09/619,252  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 60/144,722  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: 60/167,785  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 60/276,994  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/280,898  
; PRIOR FILING DATE: 2001-04-02  
; PRIOR APPLICATION NUMBER: 60/332,241  
; PRIOR FILING DATE: 2001-11-14  
; PRIOR APPLICATION NUMBER: 60/288,062  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: 60/291,766  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: 60/314,007  
; PRIOR FILING DATE: 2001-08-21  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 40  
; LENGTH: 4810  
; TYPE: DNA  
; ORGANISM: human  
US-10-074-566-40

Query Match 17.4%; Score 73.8; DB 15; Length 4810;  
Best Local Similarity 60.9%; Pred. No. 1.6e-14;  
Matches 120; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
QY 159 AGGAGACTGTGTCCCGGAGGAGTCCAGAGTCCGCGAGTGTGTTCATGTTGACGG 218  
DB 23 AGATGAATGTGTCCCGGAGGCATCCAAACGAGGGCTGTGTGCTCATGTGGAGGG 82  
QY 219 GTGGACAAGTCACTGTCTAACTGTGTGAGAGCAACAGGCCTCCAAAGGAAGAAGTTG 278



QY 225 AAGTCACCTGTCTACTGTGTGAGCAACAGCGCTCCAAAGGAAGAAGTCTGTTTCCG 284  
Db 74786 TGGGACACCTGTGTGAGCGTTTGGGAAGAGGGCTTGGGGCCCTGTGGAGCGTGCC 74727  
QY 285 AGTTTGACCTGGCAGCAGTACCTCTTTTCACTGGGAGGTTTCTGACTGGCACCAGTGTGT 344  
Db 74726 CAGGGGTGACTGCCACGGTGTCTTCTTTCCAGGGAACCTTCTGACCAGCGCTCATGGGCT 74667  
QY 345 GCT 347  
Db 74666 CCT 74664

RESULT 15  
US-10-292-798-1095/c  
Sequence 1095, Application US/10292798  
Publication No. US2003023583A1  
GENERAL INFORMATION:  
APPLICANT: SUWA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ABURATANI, HIROYUKI  
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
FILE REFERENCE: 084335/166  
CURRENT APPLICATION NUMBER: US/10/292,798  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 10/017,161  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: JP 2001-246789  
PRIOR FILING DATE: 2001-06-18  
NUMBER OF SEQ ID NOS: 2070  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1095  
LENGTH: 358246  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
LOCATION: source  
FEATURE:  
LOCATION: (1)...(358246)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (201)...(326)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (605)...(1294)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (29165)...(29206)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (39051)...(39221)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (72539)...(72763)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (133049)...(133353)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (261460)...(261526)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (291173)...(291325)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (291513)...(291770)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (302538)...(302800)  
FEATURE:  
NAME/KEY: CDS

LOCATION: (325058)...(325316)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (330441)...(330609)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (337750)...(337828)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (343822)...(344071)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (358008)...(358046)  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (547)...(546)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (1659)...(1758)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (2818)...(2917)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (17179)...(17278)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (17497)...(17497)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (17499)...(17499)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (17518)...(17518)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (51382)...(51481)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (64360)...(64459)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (71021)...(71120)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (77979)...(78078)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (81210)...(81309)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (103441)...(103540)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (130963)...(130964)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (130976)...(130977)  
OTHER INFORMATION: a, t, c, g, unknown or other

FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (130979)..(130979)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (130984)..(130984)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (130987)..(130987)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (130997)..(130999)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131002)..(131002)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131019)..(131020)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131022)..(131022)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131025)..(131025)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131031)..(131031)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131034)..(131034)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131048)..(131049)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131059)..(131060)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131063)..(131063)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131065)..(131065)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131070)..(131070)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131090)..(131091)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (131127)..(131129)  
OTHER INFORMATION: a, t, c, g, unknown or other  
FEATURE:  
NAME/KEY: modified\_base

Query Match 8.1%; Score 34.2; DB 15; Length 358246;  
Best Local Similarity 49.2%; Pred. No. 6.3;

Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
QY 77 TCTTCTCTCCATGCAGCTCATTTGGAGGCAAAAGGATAATCAGTTCTATCTGGAAC 136  
Db 352335 TATGGATGCTCTAGACAGCCCAATGTAGCAATTTCCCCAGGATGCAGTGTGCCACTTCAGCT 352276  
QY 137 CAGGTCCGTGGGGAAGGTGTACAGGAGACTGTGTCCCGGAGGAGTCCAGAGTCCGGGCG 196  
Db 352275 CAGGCTCGAGGGGCGAGGCGACAGTAGCAACTGGAGTGGTAGATGAAGTASTTCTACAG 352216  
QY 197 TGTGTGTTTTTCATGTTGACGGGTGGCAAGTCACTGTCTTAACCTGTGTGAGAGCAACA 256  
Db 352215 TAGCATGCCCGCATGGGGTAGGGTGTAGCAGAGCTCTTCTCAGGGATGGCAGACCACCA 352156  
QY 257 GGC 259  
Db 352155 GGC 352153  
Search completed: February 24, 2004, 00:59:52  
Job time : 149.513 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

XM protein - protein search, using sw model

run on: February 24, 2004, 00:32:36 ; Search time 4.94835 Seconds  
(without alignments)  
2760.353 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MPPKSNLTVTCWWSMRKL.....EVKPTAEVTAQHGLQHRM 142

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	119.5	14.4	951	2 T00260	hypothetical prote
2	117.5	14.2	898	2 T14764	hypothetical prote
3	114.5	13.8	1059	2 T22545	hypothetical prote
4	114	13.7	951	2 T00017	gene ADAMTS-1 prot
5	109.5	13.2	2165	2 T21371	hypothetical prote
6	107.5	13.0	1205	2 T18517	procollagen N-endo
7	103	12.4	1588	2 C89114	protein C37C3.6a f
8	103	12.4	2167	2 T34395	hypothetical prote
9	92	11.1	550	2 T47158	hypothetical prote
10	90.5	10.9	111	2 T34565	hypothetical prote
11	88	10.6	1184	2 T09484	cartilage intermed
12	85	10.2	142	2 JT0573	retinoic acid-indu
13	82	9.9	1444	2 T18856	angiogenesis inhib
14	81.5	9.8	1497	2 T49607	procollagen type V
15	79.5	9.6	788	2 T25061	hypothetical prote
16	79.5	9.6	957	2 T15976	hypothetical prote
17	79.5	9.6	1584	2 T00026	brain-specific ang
18	76.5	9.2	143	2 JH0385	midkine precursor
19	76.5	9.2	837	2 T00355	hypothetical prote
20	76	9.2	380	2 G01639	transmembrane prot
21	75.5	9.1	497	2 I38966	cytochrome P450 -
22	75.5	9.1	974	2 T04910	hypothetical prote
23	74	8.9	1486	1 CGH02V	collagen alpha 2IV
24	73.5	8.9	1074	2 JC5928	semaphorin F precu
25	73.5	8.9	1344	2 S47412	gene P2 protein -
26	72.5	8.7	232	2 S23398	serine proteinase
27	72.5	8.7	4544	1 S02392	alpha-2-macroglobu
28	71.5	8.6	443	2 T01697	omega-3 fatty acid
29	71.5	8.6	1548	2 S34583	serine proteinase

agglutinin precurs  
hypothetical prote  
brain-specific ang  
hypothetical 82K p  
hypothetical 82K p  
omega-3 fatty acid  
dynein gamma heavy  
hypothetical 41.9K  
hypothetical prote  
hypothetical prote  
hypothetical prote  
env polyprotein -  
coagulation factor  
capsaicin receptor  
hypothetical prote  
ferredoxin [2Fe-2S  
VP5 protein - porc

#### ALIGNMENTS

##### RESULT 1

T00260  
hypothetical protein KIAA0605 - human  
C:Species: Homo sapiens (man)  
C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 21-Jul-2000  
C:Accession: T00260  
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.  
DNA Res. 5, 31-39, 1998  
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complet  
A:Reference number: Z14086; MUID:98290545; PMID:9628591  
A:Accession: T00260  
A>Status: preliminary; translated from GE/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-951 <NAG>  
A:Cross-references: EMBL:AB011177; NID:G3043733; PIDN:BAA25531.1; PID:G3043734  
A:Experimental source: brain  
C:Genetics:  
A:Note: KIAA0605  
C:Superfamily: thrombospondin type 1 repeat homology  
F:46-106/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 14.4%; Score 119.5; DB 2; Length 951;  
Best Local Similarity 34.1%; Pred. No. 0.00078;  
Matches 30; Conservative 8; Mismatches 39; Indels 11; Gaps 4;  
QY 31 AAHLEGGKDNQFTWKPGWGRCTGDCGPGVQSRVWCFHVG---WTSLSNCGSNRP 87  
DB 785 AIHPCGDKNCPAHWLAQDWERCNTTCG-RGVKKRLVLCMELANGKPTQTRSGPGCLAKP 843  
QY 88 PKERSCF-RVCDWHSDDLFOVEVSDWHHC 114  
DB 844 PEESTCFERPC-----FKWYTSFWSEC 865

##### RESULT 2

T14764  
hypothetical protein DKFZp434H204.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T14764  
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: Z18181  
A:Accession: T14764  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-898 <WAM>  
A:Cross-references: EMBL:AL110226  
A:Experimental source: adult testis; clone DKFZp434H204  
C:Genetics:  
A:Note: DKFZp434H204.1

Query Match 14.2%; Score 117.5; DB 2; Length 898;  
Best Local Similarity 29.8%; Pred. No. 0.0012;  
Matches 31; Conservative 4; Mismatches 42; Indels 27; Gaps 4;

QY 44 WKPGWGRCTGCGGGVQSVRAVWCFHVDGWTSHLS-----NCGESNRPKRSCTFR--VC 97  
DB 95 WWAGWQICSSCGFGGLSRAVLICRSVGLDEQSALEPPACEHLPRPTTETPCNRHVP 154  
QY 98 DWHSDFQWEYSVDWHHCVLVFPYARGEVKPRTAEVTAQHGLQHR 141  
DB 155 P-----ATWAVGNWSQC-----SVTCGEGTQRR 177

RESULT 3  
T22545  
hypothetical protein F53B6.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22545  
R:White, S.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19578  
A:Accession: T22545  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1059 <WIL>  
A:Cross-references: EMBL:Z81086; PIDN:CAB03121.1; GSPDB:GN00019; CESP:F53B6.2  
A:Experimental source: clone F53B6  
C:Genetics:  
A:Gene: CESP:F53B6.2  
A:Map position: 1  
A:Introns: 38/3; 92/3; 131/3; 159/3; 236/2; 282/3; 349/2; 376/2; 420/1; 576/3; 759/3; 84

Query Match 13.8%; Score 114.5; DB 2; Length 1059;  
Best Local Similarity 28.6%; Pred. No. 0.0027;  
Matches 34; Conservative 10; Mismatches 32; Indels 43; Gaps 8;

QY 44 WKPGWGRCTGCGGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKRSCTFR----- 94  
DB 296 WEAGKSKCTASCG-GVRRHVAC--VGG-----SDCDEGRPRQETTCVAGIPCSAT 347

QY 95 RVCDW-----HSDLFOWE-----VSDWHCVLPYARGEVKPRTAEVCV 132  
DB 348 NSLDWNDRAYLDGNTFGSMNDHND---WQAPRLVAGENSTCSST--CGTGVMSRTVECV 401

RESULT 4  
T00017  
gene ADAMTS-1 protein - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
C:Accession: T00017  
R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.  
Genomics 46, 466-471, 1997  
A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene  
A:Reference number: Z14055; MUID:98110583; PMID:9441751  
A:Accession: T00017  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-951 <KUN>  
A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BA24501.1; PID:g2809057  
A:Experimental source: strain 129SVJ  
C:Genetics:  
A:Gene: ADAMTS-1  
A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2  
C:Superfamily: thrombospondin type 1 repeat homology  
F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 13.7%; Score 114; DB 2; Length 951;  
Best Local Similarity 28.7%; Pred. No. 0.0027;  
Matches 37; Conservative 15; Mismatches 49; Indels 28; Gaps 8;

QY 26 LLLSHA-----AHLEGGKQNF-----TWKPGPWGRCTGCGGGVQSVRAVWCFHV 71

DB 810 LVVGHALRPKIKFTYFMKKKTESFNAITPSEWTEENGESKTCG-SGWQRRVVQCRDI 868  
QY 72 DQWTSHLNCGESNRPKRSCTFRVCDWHSDL--PQWEVSDWHHCVLVYPYARGEVKPRTA 129  
DB 869 NGHPA--SECAKEVPKASTRPC-----ADLPCHWQVGDWSPCKT-CGKG-YKKRTL 917  
QY 130 ECVTAQHGL 138  
DB 918 KCVSHDGGV 926

RESULT 5  
T21371  
hypothetical protein F25H8.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T21371; T24896  
R:Gajadaty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19413  
A:Accession: T21371  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <WIL>  
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3  
A:Experimental source: clone F25H8  
R:Gajadaty, S.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19949  
A:Accession: T24896  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2165 <WIL>  
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3  
A:Experimental source: clone T13H10  
C:Genetics:  
A:Gene: CESP:F25H8.3  
A:Map position: 4  
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 81

Query Match 13.2%; Score 109.5; DB 2; Length 2165;  
Best Local Similarity 28.6%; Pred. No. 0.015;  
Matches 28; Conservative 13; Mismatches 32; Indels 25; Gaps 4;

QY 1 MPKSNLTCTVQWVRSRMLFLLLSLLSHAHLSGKXDNQFIWKPGWGRCTGCGGG 60  
DB 1865 LFPKPNESQTCBL-----NPDSEFPKWSFGPWGECSSKCGQ-G 1901

QY 61 VQSRVWCFHVDGWTSHLSNCGESNRPKRSCTFRVC 97  
DB 1902 IRRRRVKVANDGRVVRVKC-TTKKPRTOYCFERN 1938

RESULT 6  
T18517  
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine  
N:Alternate names: procollagen N-proteinase  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18517  
R:Collige, A.; Nusgens, B.V.; Lapiere, C.M.  
submitted to the EMBL Data Library, February 1996  
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.  
A:Reference number: Z18941  
A:Accession: T18517  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1205 <COL>  
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1  
A:Experimental source: skin  
C:Genetics:  
A:Gene: PC I-NP

**THE UNIVERSITY OF CHICAGO**

```
Db 11 WFTSPWSPGSRSC-QGGTQTRVQCLSTNQTLS--TRCPPQLRPSKRKPCNSQPCSQRPD 67
QY 103 LFQWEVSDWHCVLPVYARGEVK-RTAEC 131
Db 68 -DQCKDSS-PHCPLVQARLCVYPYTATC 95

RESULT 11
T09484
cartilage intermediate layer protein precursor - human
C:Species: Homo sapiens (man)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T09484
R:Lorenzo, P.; Neame, P.; Sommarin, Y.; Heinigard, D.
J. Biol. Chem. 273, 23469-23475, 1998
A:Title: Cloning and deduced amino acid sequence of a novel cartilage protein (CILP) ide
A:Reference number: 216689; MUID:98389785; PMID:9722584
A:Accession: T09484
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1184 <LOR>
A:Cross-references: EMBL:AF035408; NID:G3513502; PIDN:AAC33838.1; PID:G3513503
A:Experimental source: tissue type articular cartilage
C:Genetics:
A:Note: CILP
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1184/Product: cartilage intermediate layer protein #status predicted <NAT>

Query Match 10.6%; Score 88; DB 2; Length 1184;
Best Local Similarity 34.0%; Pred. No. 1.1;
Matches 17; Conservative 9; Mismatches 18; Indels 6; Gaps 2;

QY 37 KKNQFTWKP-GPMGRCTGCGPGGVQSVRAVCFPHVDGWTSHLSNGESN 85
Db 145 RRDTERIWSPWSPKSGSACGGQTGQTRICL-----AEWVSLCSEAS 189

RESULT 12
JT0573
retinoic acid-induced heparin-binding protein precursor - chicken
A:Alternate names: RIHB protein
C:Species: Gallus gallus (chicken)
C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C:Accession: JT0573; PT0032; S48199; S39266
R:Urios, P.; Duprez, D.; Le Caer, J.P.; Courtois, Y.; Vigny, M.; Laurent, M.
Biochem. Biophys. Res. Commun. 175, 617-624, 1991
A:Title: Molecular cloning of Ri-HB, a heparin binding protein regulated by retinoic aci
A:Reference number: JT0573; MUID:91207359; PMID:2018506
A:Accession: JT0573
A:Molecule type: mRNA
A:Residues: 1-142 <URI>
A:Experimental source: embryo
R:Raulais, D.; Lagente-Chevallier, O.; Guettet, C.; Duprez, D.; Courtois, Y.; Vigny, M.
Biochem. Biophys. Res. Commun. 174, 708-715, 1991
A:Title: A new heparin binding protein regulated by retinoic acid from chick embryo.
A:Reference number: PT0032; MUID:91128406; PMID:1993066
A:Accession: PT0032
A:Molecule type: protein
A:Residues: 22-77 <RAU>
R:Duprez, D.; Treaggar, J.; Pecqueur, C.; Vigny, M.
Eur. J. Biochem. 224, 931-941, 1994
A:Title: Organization and promoter activity of the retinoic-acid-induced-heparin-binding
A:Reference number: S48199; MUID:95010085; PMID:7925417
A:Accession: S48199
A:Molecule type: DNA
A>Status: preliminary
A:Residues: 1-88, 'R', 90-142 <DUP>
A:Cross-references: EMBL:X76482
R:Vigny, M.R.
submitted to the EMBL Data Library, December 1993
A:Description: Organization and promoter activity of the RIHB gene.
A:Reference number: S39266
A:Accession: S39266
```

```

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-88, 'R', 94-142 <DU2>
A:Cross-references: EMBL:X76482; NID:G434357; PIDN:CAA54020.1; PID:G434358
C:Genetics:
A:Introns: 27/1; 81/1; 135/1
C:Superfamily: pleiotrophin
C:Keywords: growth factor; heparin binding
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-142/Product: retinoic acid induced heparin-binding protein #status experimental <MA

Query Match 10.2%; Score 85; DB 2; Length 142;
Best Local Similarity 27.4%; Pred. No. 0.34;
Matches 37; Conservative 15; Mismatches 39; Indels 44; Gaps 8;

QY 18 RKLFLLSLLLSHAHLKGGKDNQFIKPGP-----WGRC---TGCGPGGVQSVRAVVC 68
Db 4 RGLLLLLLLLLLAAAANAANKKMKKEGSECODWHPGCIPIKSKDGLG----- 54

QY 69 FHVDTGWTSHLSNGESNRPPPKERSCFRVCWDHSLD---FQWEVSDWHCVLPVYARGEVK 125
Db 55 -YREG-----SCGDESKLK---CKIPCNWKKKFGADCKYKFSWGCG---SAKTGVK 100

QY 126 PRT-----AEC 131
Db 101 TRSGILKKALYNAEC 115

RESULT 13
T18856
angiogenesis inhibitor homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18856; T24653
R:McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19031
A:Accession: T18856
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WIL>
A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:CO2B4.1
A:Experimental source: clone CO2B4
R:McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19917
A:Accession: T24653
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WI2>
A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:CO2B4.1
C:Genetics:
A:Gene: CESP:CO2B4.1
A:Map position: X
A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3; 566

Query Match 9.9%; Score 82; DB 2; Length 1444;
Best Local Similarity 21.3%; Pred. No. 5.3;
Matches 32; Conservative 16; Mismatches 48; Indels 54; Gaps 7;

QY 42 FIWKFGPMGRCTGDCPGGVQSVRAVCFPHVDGWTSHLSNGESNRPPPKERSCFRVCWDHWS 101
Db 1069 FIW--SDWSSCKSGQDGIQTRQKLCIFNN---AECSSYAESRRCKDLPSCSSISGRT 1123

QY 102 -----DLFQW-EVSDWHHC-----VJ---VPYARGEVK 125
Db 1124 ISENGFDAPRWSEWSSWSACSFSLTSTRRFPQVVDPTVQGFCAILEIQIPCAPGSCS 1183

QY 126 PR-----TAECVTAQHGLOHRM 142
Db 1184 PSAGGWSLWSEWSSCKDGGTGHQIRNM 1213
```

```
Db 736 CEWSPCSTQLACEVGVQSRQCV 759

Search completed: February 24, 2004, 01:08:27
Job time : 6.94835 secs

RESULT 14
I49607
procollagen type V alpha 2 - mouse
S:Species: Mus musculus (house mouse)
D:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
Z:Accession: I49607
A:Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
Dev. Dyn. 195, 113-120, 1992
A:Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
A:Reference number: I49607; PMID:93214071; PMID:1297453
A:Accession: I49607
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1497 <RES>
A:Cross-references: GB:I02918; NID:G309180; PIDN:AAA37440.1; PID:G309181
C:Genetics:
A:Gene: Col5a-2
S:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:39-98/Domain: von Willebrand factor type C repeat homology <WVC>
F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 9.8%; Score 81.5; DB 2; Length 1497;
Best Local Similarity 26.2%; Pred. No. 6.1;
Matches 27; Conservative 12; Mismatches 33; Indels 31; Gaps 5;

2Y 14 WRSMRKLFLLSLLSHAHLEGGKD-----NQFIWKGPGWGRCTG 54
Db 5 WVGAREFL-LILSVLLGYCVSIAQOEENDEYDEIACTQHGMVLYNRDIWKSPQCICVC 63
2Y 55 DCGPGGVQSRVAVWCFVHDGWTSHLSNCGESNRPKERSCFRVC 97
Db 64 DNG-----AILCDKIE--CPEVLNCAPIPTPGE--CCPVC 95

RESULT 15
I25061
hypothetical protein T21B6.3 - Caenorhabditis elegans
S:Species: Caenorhabditis elegans
D:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
Z:Accession: T25061
A:Cottage, A.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19975
A:Accession: T25061
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-788 <NIL>
A:Cross-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3
A:Experimental source: clone T21B6
C:Genetics:
A:Gene: CESP:T21B6.3
A:Map position: X
A:Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 9.6%; Score 79.5; DB 2; Length 788;
Best Local Similarity 18.1%; Pred. No. 5.4;
Matches 37; Conservative 18; Mismatches 62; Indels 87; Gaps 9;

2Y 8 TVTCGWRSMRKLFLLSLLSHAHLEGGKNO-----FIWKP-GPWGRCTGDCGP 58
Db 564 SVTCGGGAVRQRTCLGGVFGDHLG--QGPKTEQRACDGGPCSLWSPWQEWSTCSACGS 621
2Y 59 GGVSRAV-----WCFHVDGWTSHLSNCGESNRPKPKR 91
Db 622 GMKRRQRCVQFGTDCQCPNEESFCYGPACAEWTEWC-----EWSGSSKCGPGQR-TRTR 676
2Y 92 SCF-----RVCDSHSLDLPQWEVSD----- 110
Db 677 GCLGPNQGEATTCCGSPSIETTLCEGSCCNW-SEWCHWSMCDKCCGGQGVRYEYFNRTG 735
2Y 111 --WHHCVLVPYARGEYKPRTAECV 132
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

Run on: February 23, 2004, 20:35:55 ; Search time 3.97635 Seconds  
(without alignments)  
1859.483 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MFPKSNLVTWVWSMRKL.....EVKPTAEVTAQHGLQHRM 142

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	143	17.2	1906	1	AT20 MOUSE
2	128.5	15.5	1095	1	AT17 HUMAN
3	127	15.3	1935	1	AT19 HUMAN
4	123.5	14.9	1210	1	AT19 MOUSE
5	121	14.6	1911	1	AT20 HUMAN
6	116	14.0	562	1	AT15 MOUSE
7	114	13.7	968	1	AT21 MOUSE
8	113.5	13.7	1213	1	AT22 MOUSE
9	112.5	13.6	1223	1	AT14 HUMAN
10	111	13.4	950	1	AT15 HUMAN
11	110.5	13.3	1207	1	AT19 HUMAN
12	110	13.3	967	1	AT21 RAT
13	109.5	13.2	1211	1	AT22 HUMAN
14	108.5	13.1	1077	1	AT10 HUMAN
15	107.5	13.0	1205	1	AT22 BOVIN
16	103	12.4	525	1	AT13 HUMAN
17	102.5	12.3	1205	1	AT13 HUMAN
18	102	12.3	1593	1	AT12 HUMAN
19	100.5	12.1	1224	1	AT16 HUMAN
20	99.5	12.0	377	1	AT16 MOUSE
21	96.5	11.6	1081	1	AT18 HUMAN
22	96	11.6	890	1	AT22 BOVIN
23	92	11.1	967	1	AT11 HUMAN
24	88.5	10.7	323	1	TNR6 BOVIN
25	88.5	10.7	450	1	AT10 MOUSE
26	85.5	10.3	930	1	AT22 MOUSE
27	85	10.2	142	1	MK_CHICK
28	81.5	9.8	630	1	AT24 RAT
29	81.5	9.8	930	1	AT25 HUMAN
30	79.5	9.6	1584	1	BAL1 HUMAN
31	77	9.3	860	1	AT26 HUMAN
32	77	9.3	905	1	AT28 MOUSE
33	76.5	9.2	143	1	MK_HUMAN

## RESULT 1

ID	AT20 MOUSE	STANDARD;	PRT;	1906 AA.
AC	P59511;			
DT	10-OCT-2003 (Rel. 42, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	ADAMTS-20 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 20) (ADAM-TS 20) (ADAM-TS20).			
GN	ADAMTS20.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=22566039; PubMed=12562771;			
RT	Llanazares M., Cal S., Quesada V., Lopez-Otin C.;			
RT	"Identification and characterization of ADAMTS-20 defines a novel subfamily of metalloproteinases-disintegrins with multiple thrombospondin-1 repeats and a unique GON domain.";			
RL	J. Biol. Chem. 278:13382-13389(2003).			
[2]				
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND DISEASE.			
RC	STRAIN=DBA/2;			
RX	MEDLINE=22806432; PubMed=12925592;			
RA	Rao C., Feenizler D., Loftus S.K., Liu S., McPherson J.D.,			
RA	Jurgers K.A., Apté S.S., Pavan W.J., Beier D.R.;			
RT	"A defect in a novel ADAMTS family member is the cause of the belted white-spotting mutation.";			
RL	Development 130:4665-4672(2003).			
CC	FUNCTION: May play a role in tissue-remodeling process occurring in both normal and pathological conditions.			
CC	COFACTOR: Binds 1 zinc ion per subunit (By similarity).			
CC	SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).			
CC	ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=2;			
CC	Name=1; Synonyms=ADAMTS20 B long isoform;			
CC	Isoid=P59511-1; Sequence=displayed;			
CC	Name=2; Synonyms=ADAMTS20 A short isoform;			
CC	Isoid=P59511-2; Sequence=VSP_007606, VSP_007607;			
CC	TISSUE SPECIFICITY: Expressed at low level in testis and brain.			
CC	PM: The precursor is cleaved by a furin endopeptidase (By similarity).			
CC	DISEASE: Defects in ADAMTS20 are the cause of the belted (bt) phenotype. It is a pigmental defect which occurs as a result of a defect in melanocyte development.			
CC	SIMILARITY: Belongs to peptidase family M12B.			
CC	SIMILARITY: Contains 1 disintegrin-like domain.			
CC	SIMILARITY: Contains 1 GON domain.			
CC	SIMILARITY: Contains 15 TSP type-1 domains.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			

075173 homo sapien  
062217 mus musculu  
077336 sus scrofa  
P05997 homo sapien  
Q13591 homo sapien  
Q9WJ19 mus musculu  
Q04592 mus musculu  
P11218 urtica dioi  
Q95116 bos taurus  
O60242 homo sapien  
P14728 xanthomonas

34 76.5 9.2 837 1 AT24 HUMAN  
35 76.5 9.2 1077 1 SM5A MOUSE  
36 76 9.2 332 1 TNR6 PIG  
37 74 8.9 1496 1 CA25 HUMAN  
38 73.5 8.9 1074 1 SM5A HUMAN  
39 72.5 8.7 4543 1 PBL4 MOUSE  
40 72.5 8.6 4544 1 LRPI HUMAN  
41 71.5 8.6 1877 1 PK55 MOUSE  
42 71 8.6 372 1 AGI URDI  
43 71 8.6 1170 1 TSP2 BOVIN  
44 71 8.6 1522 1 BAL3 HUMAN  
45 70 8.4 784 1 YAV2\_XANCV

## ALIGNMENTS

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

EMBL; AJ512753; CAD54808.3; -
EMBL; AY189815; AAO74895.1; -
EMBL; AY189816; AAO74896.1; -
MGB; MGI:260628; Adamts20.
InterPro; IPR001762; Disintegrin.
InterPro; IPR006025; Pept_M_Zn_BS.
InterPro; IPR001590; Peptidase_M12B.
InterPro; IPR002870; Peptidase_M12B_N.
InterPro; IPR000884; TSP1.
Pfam; PF01562; Pep_M12B_propep; 1.
Pfam; PF01421; Reprolysin; 1.
SMART; PF00090; TSP_1; 1.
PROSITE; PS00209; TSP_1; 14.
PROSITE; PS00215; ADAM_MEPRO; 1.
PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
PROSITE; PS00214; DISINTEGRIN_2; FALSE_NEG.
PROSITE; PS00092; TSP1; 13.
PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
FT SIGNAL 1 26
Repeat; Extracellular matrix; Alternative splicing.
FT PROPEP 27 249
FT CHAIN 250 1906
FT DOMAIN 250 464
FT DOMAIN 465 552
FT DOMAIN 553 608
FT DOMAIN 609 720
FT DOMAIN 721 842
FT DOMAIN 843 901
FT DOMAIN 906 962
FT DOMAIN 962 1015
FT DOMAIN 1017 1074
FT DOMAIN 1075 1131
FT DOMAIN 1148 1202
FT DOMAIN 1203 1260
FT DOMAIN 1300 1351
FT DOMAIN 1354 1411
FT DOMAIN 1412 1465
FT DOMAIN 1468 1526
FT DOMAIN 1527 1584
FT DOMAIN 1585 1648
FT DOMAIN 1650 1706
FT DOMAIN 1707 1906
FT METAL 399 399
FT ACT_SITE 400 400
FT METAL 403 403
FT METAL 403 403
FT CARBOHYD 92 92
FT CARBOHYD 221 221
FT CARBOHYD 714 714
FT CARBOHYD 798 798
FT CARBOHYD 805 805
FT CARBOHYD 1057 1057
FT CARBOHYD 1562 1562
FT CARBOHYD 1719 1719
FT CARBOHYD 1759 1759
FT CARBOHYD 1777 1777
FT VARSPPLIC 1424 1425
FT Missing (in isoform 2).
FT VARSPLIC 1426 1906
FT CONFLICT 1211 1211
FT CONFLICT 1262 1262
FT SEQUENCE 1906 AA; 212041 MW; 0EB2568547B557EA CRC64;

```

Query Match

17.2%; Score 143; DB 1; Length 1906;

Best Local Similarity 32.4%; Pred. No. 1.4e-06;  
Matches 35; Conservative 14; Mismatches 41; Indels 18; Gaps 6;  
QY 38 KDNQIWKPGWGRGTGCGGGVQSVRAVCFHVDGWTSHLSNCGESNRPKERSCFRVC 97  
DB 1300 RGNQ--WRTGPGWAGSRSCA-GGLQHRVAVCODEDGRSA--TSCDGGSKPPESRHC----- 1350  
QY 98 DWHSDLFWGVSDMHCHLVLPYARGEVPRPTAEC-----VTAQHGLQ 139  
DB 1351 -CGGCPCHNNYGDGECTQT--CGGGVKSFRVICQPPNGQMTQEHSC 1395

## RESULT 2

```

AT17_HUMAN
ID AT17_HUMAN STANDARD; PRT; 1095 AA.
AC Q8TE56;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAMTS-17 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 17) (ADAM-TS 17) (ADAM-TS17).
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains."
RL Gene 283:49-62(2002).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed in fetal lung, in adult brain,
CC prostate, and liver.
CC -1- PTM: The precursor is cleaved by a furin endopeptidase (By
CC similarity).
CC -1- SIMILARITY: Belongs to peptidase family M12B.
CC -1- SIMILARITY: Contains 1 disintegrin-like domain.
CC -1- SIMILARITY: Contains 1 PLAC domain.
CC -1- SIMILARITY: Contains 5 TSP type-1 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```
DR PROSITE; P50214; DISINTEGRIN_2; FALSE_NEG.
DR PROSITE; P50900; PLAC; 1.
DR PROSITE; P50092; TSPI; 5.
DR PROSITE; P50142; ZINC PROTEASE; 1.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
DR Repeat; Extracellular matrix.
DR SIGNAL; 1 27
DR PROPEP; 28 223
DR CHAIN; 224 1095
DR DOMAIN; 224 452
DR DOMAIN; 224 452
DR DOMAIN; 453 542
DR DOMAIN; 543 598
DR DOMAIN; 599 701
DR DOMAIN; 702 779
DR DOMAIN; 800 860
DR DOMAIN; 861 922
DR DOMAIN; 925 968
DR DOMAIN; 972 1029
DR DOMAIN; 1045 1084
DR DOMAIN; 1095 1201
DR SITE; 201 201
DR METAL; 389 389
DR ACT_SITE; 390 390
DR METAL; 393 393
DR METAL; 399 399
DR CARBOHYD; 167 167
DR CARBOHYD; 483 483
DR CARBOHYD; 785 785
DR CARBOHYD; 790 790
DR CARBOHYD; 832 832
DR CARBOHYD; 839 839
DR CARBOHYD; 894 894
DR SEQUENCE; 1095 AA; 121099 MW; 45654048475405D CRC64;

Query Match 15.5%; Score 128.5; DB 1; Length 1095;
Best Local Similarity 31.1%; Pred. No. 2.4e-05;
Matches 33; Conservative 16; Mismatches 40; Indels 17; Gaps 7;

Y 35 EGKKNQFTWKPGRCTGDCGGVQSGRAYWCFHVDGWISHLNCSGESNPPKRSCTP 94
b 917 EG-QDCLSIWEASESQCSASCK-GWKRTVAC-----TNSQKCDASTPRAEAA-- 966

Y 95 RYCDHMSDLFOWESDWHVCLVPYARGVEKPRTAEC---VTAQHG 137
b 967 --CEDYSGCYEYWKTDGWSCTST-CGKG-LQSRVQCKHKTGRHG 1008

RESULT 3
D AT9 HUMAN STANDARD; PRT; 1935 AA.
T Q9P2N4; Q9NR29;
T 16-OCT-2001 (Rel. 40, Created)
T 10-OCT-2003 (Rel. 42, Last sequence update)
T 10-OCT-2003 (Rel. 42, Last annotation update)
E ADAMTS-9 precursor (EC 3.4.24.4) (A disintegrin and metalloproteinase
E with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
E ADAMTS9 OR KIAA1312.
S Homo sapiens (Human).
S Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
S Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
S NCBI_TaxId=9606;
S [1]
S SEQUENCE FROM N.A. (ISOFORM 3).
S TISSUE=Fetal;
S MEDLINE=20396138; PubMed=10936055;
S Clark M.E., Kelnner G.S., Turberville L.A., Boyer A., Arden K.A.,
S Maki R.A.;
S "ADAMTS-9, a novel member of the ADAM-TS/Metalloprotein gene
S family.";
S Genomics 67:343-350(2000).
S [2]
S SEQUENCE FROM N.A. (ISOFORM 1), AND FUNCTION.
S MEDLINE=22513925; PubMed=12514189;
```

Somerville R.P., Longpre J.-M., Jungers K.A., Engle J.M., Ross M.,  
Evanko S., Micht T.N., Leduc R., Apte S.S.;  
"Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS  
subfamily related to Caenorhabditis elegans GON-1";  
J. Biol. Chem. 278:9503-9513(2003).  
[3]  
SEQUENCE OF 159-1935 FROM N.A. (ISOFORM 2).  
TISSUE=Brain;  
MEDLINE=20181126; PubMed=10718198;  
Nagase T., Kikuno R., Ishikawa K.-I., Hirokawa M., Ohara O.;  
"Prediction of the coding sequences of unidentified human genes. XVI.  
The complete sequences of 150 new cDNA clones from brain which code  
for large proteins in vitro.";  
DNA Res. 7:65-73(2000).  
CC -!- FUNCTION: Cleaves the large aggregating proteoglycans, aggrecan  
and versican.  
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 1838-Glu-Ala-1839  
site and versican at the 1428-Glu-Ala-1429 site.  
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
matrix (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=3;  
Name=1; Synonyms=ADAMTS-9B;  
Isoid-Q9P2N4-3; Sequence=Displayed;  
Name=2; Synonyms=Long;  
Isoid-Q9P2N4-1; Sequence=VSP\_007548, VSP\_007549;  
Note=May result from the retention of an intron in the cDNA  
leading to a premature stop codon;  
Name=3; Synonyms=Short;  
Isoid-Q9P2N4-2; Sequence=VSP\_005499, VSP\_005500;  
CC -!- TISSUE SPECIFICITY: Highly expressed in all fetal tissues.  
Expressed in a number of adult tissues with highest expression in  
heart, placenta and skeletal muscle.  
CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important  
for a tight interaction with the extracellular matrix (By  
similarity).  
CC -!- DOMAIN: The ancillary domains, including the TSPs domain, are  
required for specific extracellular localization and for its  
versicanase and aggrecanase activities.  
CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By  
similarity).  
CC -!- SIMILARITY: Belongs to peptidase family M12B.  
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.  
CC -!- SIMILARITY: Contains 1 GON domain.  
CC -!- SIMILARITY: Contains 15 TSP type-1 domains.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC ENBL; AF261918; AAF89106.1; --  
CC ENBL; AF488803; AAO15765.1; --  
CC ENBL; AB037733; BAA92550.1; --  
CC HSP; P15167; IATL.  
CC MEROPS; M12.021; --  
CC Genew; HGNC:13202; ADAMTS9.  
CC MIM; 605421; --  
CC GO; GO:0008237; F:metallopeptidase activity; TAS.  
CC GO; GO:0007275; P:metalloprotein catabolism; TAS.  
CC GO; GO:0006516; P:glycoprotein catabolism; TAS.  
CC InterPro; IPR001762; Disintegrin.  
CC InterPro; IPR001818; Pept\_M10A\_M12B.  
CC InterPro; IPR006025; Pept\_M\_2n\_BS.  
CC InterPro; IPR001590; Peptidase M12B.  
CC InterPro; IPR002870; Peptidase\_M12B\_N.  
CC InterPro; IPR000884; TSPI.  
CC Pfam; PF01562; Pep\_M12B\_propep; 1.  
CC Pfam; PF01421; Reprolysin; 1.



DR PFAM; PF00090; tsp.1; 11.  
 DR SMART; SM00209; TSPI; 12.  
 DR PROSITE; PS00215; ADAM MEPRO; 1.  
 DR PROSITE; PS00548; CYSTEINE SWITCH; FALSE NEG.  
 DR PROSITE; PS00422; DISINTEGRIN\_1; FALSE NEG.  
 DR PROSITE; PS00214; DISINTEGRIN\_2; FALSE NEG.  
 DR PROSITE; PS00092; TSPI; 14.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1.  
 DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix; Alternative splicing.  
 FT SIGNAL 1 18  
 FT PROPEP 19 287  
 FT CHAIN 288 1935  
 FT DOMAIN 293 499  
 FT DOMAIN 509 587  
 FT DOMAIN 588 643  
 FT DOMAIN 644 752  
 FT DOMAIN 753 877  
 FT DOMAIN 878 936  
 FT DOMAIN 939 997  
 FT DOMAIN 1049 1099  
 FT DOMAIN 1052 1109  
 FT DOMAIN 1110 1166  
 FT DOMAIN 1182 1240  
 FT DOMAIN 1241 1296  
 FT DOMAIN 1328 1379  
 FT DOMAIN 1382 1440  
 FT DOMAIN 1441 1494  
 FT DOMAIN 1497 1555  
 FT DOMAIN 1556 1611  
 FT DOMAIN 1612 1676  
 FT DOMAIN 1677 1734  
 FT DOMAIN 1735 1935  
 FT METAL 88 96  
 FT SITE 223 223  
 FT ACT SITE 434 434  
 FT METAL 435 435  
 FT METAL 438 438  
 FT METAL 444 444  
 FT CARBOHYD 112 112  
 FT CARBOHYD 135 135  
 FT CARBOHYD 271 271  
 FT CARBOHYD 749 749  
 FT CARBOHYD 840 840  
 FT CARBOHYD 1213 1213  
 FT CARBOHYD 1267 1267  
 FT CARBOHYD 1788 1788  
 FT CARBOHYD 1806 1806  
 FT VARSPLIC 1064 1072  
 FT VARSPLIC 1073 1935  
 FT VARSPLIC 1624 1629  
 FT VARSPLIC 1630 1935  
 FT CONFLICT 46 46  
 FT CONFLICT 96 96  
 FT CONFLICT 182 182  
 FT CONFLICT 367 367  
 FT CONFLICT 1117 1117  
 FT SEQUENCE 1935 AA; 216556 MW; FD3D51E88300A3C6 CRC64;  
 Query Match 15.3%; Score 127; DB 1; Length 1935;  
 Best Local Similarity 31.5%; Pred. No. 6e-05;  
 Matches 28; Conservative 15; Mismatches 34; Indels 12; Gaps 5;  
 2y 44 WKGPWGRCTGDCPGVQSRVAVCFHVDWTHSLNCGESNPPKERSC-PRVCDWHS 102  
 1332 WRTGPWGRCTGDCPGVQSRVAVCFHVDWTHSLNCGESNPPKERSC-PRVCDWHS 102  
 103 LFQWVSDWHEHCVLPVARGVEKPRFAEC 131

Db 1385 --QWYGNWGECTKL--CGGGIRTRLVVC 1409  
 RESULT 4  
 AT19 MOUSE  
 ID AT19 MOUSE STANDARD; PRT; 1210 AA.  
 AC P59509;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ADAMTS-19 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 19) (ADAM-TS 19) (ADAM-TS19).  
 GN ADAMTS19.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Ovary;  
 RX MEDLINE=22505168; PubMed=12617826;  
 RA Menke D.B., Page D.C.;  
 RT "Sexually dimorphic gene expression in the developing mouse gonad.";  
 RL Gene Expr. Patterns 2:359-367(2002).  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 CC matrix (By similarity).  
 CC -!- TISSUE SPECIFICITY: Expressed predominantly in fetal ovary, low  
 CC levels of expression is also detected in kidney, heart, skeletal  
 CC muscle, lung and testis.  
 CC -!- DEVELOPMENTAL STAGE: Expression is strongest in anterior and  
 CC ventral regions of the ovary at 12.5 and 13.5 dpc before becoming  
 CC more uniform.  
 CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By  
 CC similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M12B.  
 CC -!- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -!- SIMILARITY: Contains 5 TSP type-1 domains.  
 CC -!- SIMILARITY: Contains 1 PLAC domain.  
 CC -!- CAUTION: By homology with the human sequence, it is uncertain  
 CC whether Met-1 or Met-5 is the initiator.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; AY135183; MAM10155.1; -  
 CC MGD; MGI:2442875; Adamts19.  
 CC InterPro; IPR006586; ADAM\_cysteine.  
 CC InterPro; IPR001762; Disintegrin.  
 CC InterPro; IPR001818; Pept M10A M12B.  
 CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro; IPR001590; Peptidase\_M12B.  
 CC InterPro; IPR003870; Peptidase\_M12B\_N.  
 CC InterPro; IPR008884; TSPI.  
 CC Pfam; PF01562; Pep\_M12B\_propep; 1.  
 CC Pfam; PF01421; Repolysin; 1.  
 CC Pfam; PF00090; tsp.1; 5.  
 CC SMART; SM00608; ACR; 1.  
 CC SMART; SM00209; TSPI; 5.  
 CC PROSITE; PS0215; ADAM MEPRO; 1.  
 CC PROSITE; PS00548; CYSTEINE SWITCH; FALSE NEG.  
 CC PROSITE; PS00422; DISINTEGRIN\_1; FALSE NEG.  
 CC PROSITE; PS0214; DISINTEGRIN\_2; FALSE NEG.  
 CC PROSITE; PS00300; PLAC; 1.  
 CC PROSITE; PS00092; TSPI; 5.  
 CC PROSITE; PS00142; ZINC PROTEASE; 1.  
 CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix.

```

FT SIGNAL 1 30
FT PROPEP 31 319
FT CHAIN 320 1210
FT DOMAIN 320 548
FT DOMAIN 549 636
FT DOMAIN 637 689
FT DOMAIN 690 783
FT DOMAIN 794 917
FT DOMAIN 918 978
FT DOMAIN 979 1040
FT DOMAIN 1042 1086
FT DOMAIN 1090 1147
FT DOMAIN 1163 1202
FT SITE 297 297
FT METAL 485 485
FT ACT_SITE 486 486
FT METAL 489 489
FT METAL 495 495
FT CARBOHYD 54 54
FT CARBOHYD 263 263
FT CARBOHYD 800 800
FT CARBOHYD 910 910
FT CARBOHYD 931 931
FT CARBOHYD 952 952
FT CARBOHYD 1012 1012
SQ SEQUENCE 1210 AA; 134560 MW; 0AB812ABAB4BB7A2 CRC64;

Query Match 14.98; Score 123.5; DB 1; Length 1210;
Best Local Similarity 30.24; Pred. No. 8.5e-05;
Matches 32; Conservative 17; Mismatches 40; Indels 17; Gaps 7;

Qy 35 EGKNDNFIWKPGWCRCTGDCGGVQSVRAVWCFHVDGWTSHLSNCGSNPPKRSCTF 94
Db 1035 EG-QDQMTVWEAGWSVCSVKCGK-GVRHRTVC-----TNRKKCVLSTPREAD-- 1084

Qy 95 RVDWHSDLQFQWEVSDVHHVCLVYPYARGEVKPRTAGC---VTAQHG 137
Db 1085 --CEDYSKYVWRVGDWSKC-SITCGKG-MQSRVIQMKHITGRHG 1126

RESULT 5
AT20_HUMAN
ID AT20_HUMAN STANDARD; PRT; 1911 AA.
AC P59510;
DC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=22513925; PubMed=12514189;
RA Somerville R.P., Longpre J.-M., Jungers K.A., Engle J.M., Ross M.,
RT Evanko S., Wight T.N., Leduc R., Apté S.S.;
RT "Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS
RL subfamily related to Caenorhabditis elegans GON-1.";
RL J. Biol. Chem. 278:9503-9513(2003).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX TISSUE=Liver;
RA Llamazares M., Cal S., Quesada V., Lopez-Otin C.;
RT "Identification and characterization of ADAMTS-20 defines a novel
RT subfamily of metalloproteinases-disintegrins with multiple
RT thrombospondin-1 repeats and a unique GON domain.";
RL J. Biol. Chem. 278:13382-13389(2003).
RL -!- FUNCTION: May play a role in tissue-remodeling process occurring.

in both normal and pathological conditions.
-!- COPACTOR: Binds 1 zinc ion per subunit (By similarity).
-!- SURCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
IsoId=P59510-1; Sequence=Displayed;
Name=2;
IsoId=P59510-2; Sequence=VSP 007106, VSP 007107, VSP 007108;
-!- TISSUE SPECIFICITY: Very sparingly expressed, although is detected
at low levels in testis, prostate, ovary, heart, placenta, lung
and pancreas. Overexpressed in several brain, colon and breast
carcinomas.
-!- PTM: The precursor is cleaved by a furin endopeptidase (By
similarity).
-!- SIMILARITY: Belongs to peptidase family M12B.
-!- SIMILARITY: Contains 1 disintegrin-like domain.
-!- SIMILARITY: Contains 1 GON domain.
-!- SIMILARITY: Contains 15 TSP type-1 domains.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
EMBL; AF488804; AAO15766.1; -
EMBL; AJ515153; CAD56159.3; -
EMBL; AJ515154; CAD56160.2; -
Genew; HGNC:17178; ADAMTS20.
InterPro; IPR001762; Disintegrin.
InterPro; IPR006025; Pept_M_Zn_BS.
InterPro; IPR001590; Peptidase_M12B.
InterPro; IPR002870; Peptidase_M12B_N.
InterPro; IPR000884; TSP1.
Pfam; PF01562; Pep_M12B_propep; 1.
Pfam; PF01421; Repolysin; 1.
Pfam; PF00090; Tsp_1; 11.
SMART; SMO0209; TSP1; 12.
PROSITE; PS0215; ADAM_MEPRO; 1.
PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
PROSITE; PS0214; DISINTEGRIN_2; FALSE_NEG.
PROSITE; PS50092; TSP1; 12.
PROSITE; PS0142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 21
FT PROPEP 22 253
FT CHAIN 254 1911
FT DOMAIN 254 467
FT DOMAIN 468 555
FT DOMAIN 556 611
FT DOMAIN 612 723
FT DOMAIN 724 846
FT DOMAIN 847 905
FT DOMAIN 906 962
FT DOMAIN 967 1024
FT DOMAIN 1025 1074
FT DOMAIN 1077 1136
FT DOMAIN 1153 1207
FT DOMAIN 1208 1265
FT DOMAIN 1305 1357
FT DOMAIN 1359 1417
FT DOMAIN 1418 1476
FT DOMAIN 1477 1532
FT DOMAIN 1536 1589
FT DOMAIN 1590 1653
FT DOMAIN 1655 1711
FT DOMAIN 1712 1911
FT METAL 403 403
ZINC (CATALYTIC) (BY SIMILARITY).

```



Kuno K., Terashima Y., Matsushima K.;  
"ADAMTS-1 is an active metalloproteinase associated with the  
extracellular matrix.";  
J. Biol. Chem. 274:18821-18826(1999).  
[5]  
FUNCTION  
MEDLINE=20389568; PubMed=10930576;  
Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,  
Ohno H., Matsushima K.;  
"ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";  
FEBS Lett. 478:241-245(2000).  
[6]  
FUNCTION, AND INDUCTION.  
MEDLINE=20243757; PubMed=10781075;  
Rooker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,  
Richards J.S.;  
"Progesterone-regulated genes in the ovulation process: ADAMTS-1 and  
cathepsin L proteases.";  
Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).  
-!- FUNCTION: Cleaves aggrecan, a cartilage proteoglycan, and may be  
involved in its turnover. Has angiogenic inhibitor activity (By  
similarity). Active metalloproteinase, which may be associated with  
various inflammatory processes as well as development of cancer  
cachexia. May play a critical role in follicular rupture (By  
similarity).  
-!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 1691-Glu-|-Leu-1692  
site, within the chondroitin sulfate attachment domain.  
-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
matrix.  
-!- INDUCTION: Induced in vitro in colon adenocarcinoma cells by  
interleukin-1, or in vivo in kidney and heart by  
lipopolysaccharide. Also induced by LH stimulation in granulosa  
cells of preovulatory follicles.  
-!- DOMAIN: The spacer domain and the TSP type-1 domains are important  
for a tight interaction with the extracellular matrix.  
-!- PTM: The precursor is cleaved by a furin endopeptidase.  
-!- SIMILARITY: Belongs to peptidase family M12B.  
-!- SIMILARITY: Contains 1 disintegrin-like domain.  
-!- SIMILARITY: Contains 3 TSP type-1 domains.  
-!- CAUTION: Ref.2 sequence differs from that shown due to a  
frameshift in position 7.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AB001735; BAA24501.1; ALT INIT.  
EMBL; D67076; BAA11088.1; ALT FRAME.  
EMBL; BC040382; AAH40382.1; -  
EMBL; BC050834; AAH50834.1; -  
MEROPS; M12.222; -  
MGP; MGI:109249; Adamts1.  
InterPro; IPR006586; ADAM cysteine.  
InterPro; IPR001762; Disintegrin.  
InterPro; IPR006025; Pept M.Zn.BS.  
InterPro; IPR001590; Peptidase\_M12B.  
InterPro; IPR002870; Peptidase\_M12B\_N.  
InterPro; IPR000884; TSP1.  
InterPro; IPR008085; TSP1.  
Pfam; PF01562; Pep M12B propep; 1.  
Pfam; PF01421; Reptolysin; 1.  
Pfam; PF00030; TSD; 1; 3.  
PRINTS; PR01705; TSP1REPEAT.  
SMART; SMC0608; ACR; 1.  
SMART; SMC0209; TSP1; 3.  
PROSITE; PS50215; ADAM MEPRO; 1.  
PROSITE; PS04027; DISINTEGRIN 1; FALSE NEG.  
PROSITE; PS50214; DISINTEGRIN 2; FALSE NEG.

DR PROSITE; PS00092; TSP1; 3;  
 DR PROSITE; PS00142; ZINC PROTEASE; 1;  
 KW Hydrolyase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix; Heparin-binding;  
 FT SIGNAL 1 48  
 FT PROPEP 49 253  
 FT CHAIN 254 968  
 FT DOMAIN 254 476  
 FT DOMAIN 477 559  
 FT DOMAIN 560 615  
 FT DOMAIN 618 725  
 FT DOMAIN 726 850  
 FT DOMAIN 855 911  
 FT DOMAIN 912 968  
 FT DOMAIN 195 159  
 FT SITE 206 206  
 FT METAL 402 402  
 FT ACT SITE 403 403  
 FT METAL 406 406  
 FT METAL 412 412  
 FT CARBOHYD 548 548  
 FT CARBOHYD 721 721  
 FT CARBOHYD 765 765  
 FT CARBOHYD 783 783  
 FT CARBOHYD 946 946  
 FT MUTAGEN 403 403  
 FT CONFLICT 335 335  
 FT CONFLICT 425 425  
 FT SEQUENCE 968 AA; 42EBDA55499FB6C1 CRC64;  
 Query Match 13.7%; Score 114; DB 1; Length 968;  
 Best Local Similarity 28.7%; Pred. No. 0.00064;  
 Matches 37; Conservative 15; Mismatches 49; Indels 28; Gaps 8;  
 QY 26 LLLSRA-----AHLEGGKQNF-----IWKEGPGWRCCTGCGPGVGSRAVWCVHV 71  
 DB 827 LMGVHALRPKIKFTFMKKKTSFNAIPTFSEWVEEGECSKTCG-SGWRVVCQCDI 885  
 QY 72 DQWTHSLSCGNSPPPKERSCFRCDWHSDL--FQWESDWHVCLVYPYARGVEKPRTA 129  
 DB 886 NQHFA--SECAKEVFPASTRPC-----ADLPCHWQVGDWSPCKT-CGKG-YKRTL 934  
 QY 130 ECVTAQHL 138  
 DB 935 KCVSHDGGV 943

## RESULT 8

AT52\_MOUSE  
 ID AT52\_MOUSE STANDARD; PRT; 1213 AA.  
 AC Q8C9W3;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ADAMTS-2 precursor (SC 3.4.24.14) (A disintegrin and metalloproteinase  
 DE with thrombospondin motifs 2) (ADAM-TS2) (Procollagen I/II  
 DE amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC  
 DE I-NP) (Procollagen N-endopeptidase) (pNPI).  
 GN ADAMTS2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; Tissue=Thymus;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Otsu N., Saito K., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Ragi K., Tomaru Y., Hasegawa Y., Nogi A., Schonbach C., Gojobori T.,  
 RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,

Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hiroxane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RN Nature 420:563-573(2002).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; Tissue=Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Sapotnick M., Soares M.B., Donaldson M.P., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Bask S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Rosak S.A., McQuinn P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Morley D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Rahay J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -|- FUNCTION: Cleaves the propeptides of type I and II collagen prior  
 CC to fibril assembly. Does not act on type III collagen. May also  
 CC play a role in development that is independent of its role in  
 CC collagen biosynthesis (By similarity).  
 CC -|- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
 CC alpha-1(I) at Pro-|-Gln and of alpha-1(II) and alpha-2(I) chains  
 CC at Ala-|-Gln.  
 CC -|- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -|- SUBUNIT: May belong to a multimeric complex. Binds specifically to  
 CC collagen type XIV (By similarity).  
 CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 CC matrix (By similarity).  
 CC -|- DOMAIN: The spacer domain and the TSP type-1 domains are important  
 CC for a tight interaction with the extracellular matrix.  
 CC -|- PFM: The precursor is cleaved by a furin endopeptidase (By  
 CC similarity).  
 CC -|- SIMILARITY: Belongs to peptidase family M12B.  
 CC -|- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -|- SIMILARITY: Contains 1 PLAC domain.  
 CC -|- SIMILARITY: Contains 4 TSP type-1 domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
-----
EMBL: AK040370; BAC30572.1; -
EMBL: BC046456; RAH46456.1; -
GCD: MGI:1347356; Adamts2.
InterPro: IPR001782; Disintegrin.
InterPro: IPR006025; Pept M Zn BS.
InterPro: IPR001590; Peptidase_M12B.
InterPro: IPR002870; Peptidase_M12B_N.
InterPro: IPR000884; TSP1.
Pfam: PF01562; Pep_M12B_propep; 1.
Pfam: PF01421; Repolysin; 1.
Pfam: PF00900; TSP1; 4.
SMART: SMO0209; TSP1; 4.
PROSITE: PS00215; ADAM_MEPPO; 1.
PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.
PROSITE: PS0214; DISINTEGRIN_2; FALSE NEG.
PROSITE: PS0900; PLAC; 1.
PROSITE: PS0092; TSP1; 4.
PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.
Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
Repeat; Collagen degradation; Extracellular matrix.
SIGNAL 1 28 POTENTIAL.
PROPEP 29 260 BY SIMILARITY.
CHAIN 29 1213 ADAMTS-2.
DOMAIN 261 471 METALLOPROTEASE.
DOMAIN 480 560 DISINTEGRIN-LIKE.
DOMAIN 561 617 TSP TYPE-1 1.
DOMAIN 618 722 CYS-RICH.
DOMAIN 723 851 SPACER.
DOMAIN 855 913 TSP TYPE-1 2.
DOMAIN 913 975 TSP TYPE-1 3.
DOMAIN 976 1030 TSP TYPE-1 4.
DOMAIN 1060 1098 PLAC.
SITE 692 694 CELL ATTACHMENT SITE (POTENTIAL).
METAL 409 409 ZINC (CATALYTIC) (BY SIMILARITY).
T ACT SITE 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
T METAL 413 413 ZINC (CATALYTIC) (BY SIMILARITY).
T CARBOHYD 111 111 N-LINKED (GLCNAC. .) (POTENTIAL).
T CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
T CARBOHYD 949 949 N-LINKED (GLCNAC. .) (POTENTIAL).
T CARBOHYD 950 950 N-LINKED (GLCNAC. .) (POTENTIAL).
T CARBOHYD 994 994 N-LINKED (GLCNAC. .) (POTENTIAL).
T CARBOHYD 1032 1032 N-LINKED (GLCNAC. .) (POTENTIAL).
T CARBOHYD 1099 1099 N-LINKED (GLCNAC. .) (POTENTIAL).
T CARBOHYD 1147 1147 N-LINKED (GLCNAC. .) (POTENTIAL).
T CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE 1213 AA; 135298 MW; B27431E00443EDB5 CRC64;

Query Match 13.7%; Score 113.5; DB 1; Length 1213;
Best Local Similarity 29.0%; Pred. No. 0.0009;
Matches 31; Conservative 12; Mismatches 37; Indels 27; Gaps 6;

40 NQFIKPGFWGRTGCGGGVQSRVAVWC---FHYDGMTS-HLSCGBSNRPKPKRSQFR 95
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
915 SQPVVVTGWEPTQSCGRTGQVRSVRCIQPLHNNTTRESVTHKCDN-HRPSERRACNR 973
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
96 -VCDWHSDFQWEVSDWHCVLPYARGVKPRTAEVTAQGLQHR 141
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
974 ELCPG-----RWRAGWSQC-----SVTCGNGTQER 999
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
UT14_HUMAN
ID AT14_HUMAN STANDARD; PRT; 1223 AA.
C Q8WXS8; Q8TE55; Q8TEY8;
D 28-FEB-2003 (Rel. 41, Created)
E 28-FEB-2003 (Rel. 41, Last sequence update)
F 15-MAR-2004 (Rel. 43, Last annotation update)
G ADAMTS-14 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
H with thrombospondin motifs 14) (ADAM-TS 14) (ADAM-TS14).
```

```
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A).
RX MEDLINE=21638061; PubMed=11779638;
RA Bolz H., Ramirez A., von Brederlow B., Kubisch C.;
RT "Characterization of ADAMTS14, a novel member of the ADAMTS
RT metalloproteinase family";
RL Biochim. Biophys. Acta 1522:221-225 (2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Fetal lung;
RX MEDLINE=21856482; PubMed=11857212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62 (2002).
RN [3]
RP SEQUENCE OF 29-1223 FROM N.A. (ISOFORMS B; C AND D), AND
RP ALTERNATIVE PROMOTER USAGE.
RX MEDLINE=21839041; PubMed=11741898;
RA Collige A., Vandenbergh I., Thiry M., Lambert C.A., Van Beeumen J.,
RA Li S.-W., Prockop D.J., Lapiere C.M., Nussgens B.V.;
RT "Cloning and characterization of ADAMTS-14, a novel ADAMTS displaying
RT high homology with ADAMTS-2 and ADAMTS-3.";
RL J. Biol. Chem. 277:5756-5766 (2002).
CC -!- FUNCTION: Has a aminoprocollagen type I activity processing
CC activity in the absence of ADAMTS2. Seems to be synthesized as a
CC latent enzyme that requires activation to display aminoprocollagen
CC peptidase activity.
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative promoter;
CC Comment=2 isoforms, A (shown here) and B, are produced by use of
CC alternative promoters;
CC Event=Alternative splicing; Named isoforms=4;
CC Name=A;
CC IsoId=Q8WXS8-1; Sequence=Displayed;
CC Name=B;
CC IsoId=Q8WXS8-2; Sequence=VSP_006958;
CC Name=C;
CC IsoId=Q8WXS8-3; Sequence=VSP_006958, VSP_005501;
CC Note=Produced by alternative splicing of isoform B;
CC Name=D;
CC IsoId=Q8WXS8-4; Sequence=VSP_005501;
CC Note=Produced by alternative splicing of isoform A;
CC -!- TISSUE SPECIFICITY: Expressed in retina and at low levels in
CC brain, lung and placenta. High expression in fetal tissues.
CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important
CC for a tight interaction with the extracellular matrix (By
CC similarity).
CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By
CC similarity).
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.
CC -!- SIMILARITY: Contains 1 PLAC domain.
CC -!- SIMILARITY: Contains 4 TSP type-1 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL; AF358666; AAL40229.1; -

```

DR EMBL; AJ345098; CAC87943.1; -.
DR EMBL; AF366351; AAL79814.1; -.
DR MEROPS; M12.024; -.
DR Genew; HGNC:14899; ADAMTS14.
DR MIM; 607506; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001818; Pept_M10A_M12B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001590; Peptidase_M12B.
DR InterPro; IPR002870; Peptidase_M12B_N.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000885; TSP1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF04421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 4.
DR PRINTS; PR01705; TSP1REPEAT.
DR SMART; SM00209; TSP1; 4.
DR PROSITE; PS02015; ADAM_MEPRO; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS02014; DISINTEGRIN_2; FALSE_NEG.
DR PROSITE; PS00900; PLAC; 1.
DR PROSITE; PS00092; TSP1; 4.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Collagen degradation; Repeat; Extracellular matrix;
KW Alternative splicing; Alternative promoter usage.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 252 BY SIMILARITY.
FT CHAIN 253 1223 ADAMTS-14.
FT DOMAIN 253 460 METALLOPROTEASE.
FT DOMAIN 461 551 DISINTEGRIN-LIKE.
FT DOMAIN 552 607 TSP TYPE-1.
FT DOMAIN 608 729 CIS-RICH.
FT DOMAIN 730 846 SPACER.
FT DOMAIN 847 907 TSP TYPE-1 2.
FT DOMAIN 908 967 TSP TYPE-1 3.
FT DOMAIN 968 1022 TSP TYPE-1 4.
FT DOMAIN 1059 1097 PLAC.
FT DOMAIN 875 878 POLY-ARG.
FT DOMAIN 1100 1221 PRO-RICH.
FT ACT_SITE 398 399 BY SIMILARITY.
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 941 941 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1027 1027 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 1 68 Missing (in isoform B and isoform C).
FT VARSPPLIC 1 68 /FTid_VSP_006958.
FT VARSPPLIC 368 368 G -> GMOG (in isoform C and isoform D).
FT CONFLICT 868 868 /FTid_VSP_005501.
FT CONFLICT 884 884 Q -> R (IN REF. 2).
FT CONFLICT 901 901 Q -> H (IN REF. 2).
FT CONFLICT 923 923 C -> S (IN REF. 2).
FT CONFLICT 1024 1024 C -> Y (IN REF. 2).
FT CONFLICT 1024 1024 N -> S (IN REF. 2).
SQ SEQUENCE 1223 AA; 133871 MW; 36394AC4D92F170F CRC64;

Query Match 13.6%; Score 112.5; DB 1; Length 1223;
Best Local Similarity 30.1%; Pred. No. 0.0011;
Matches 34; Conservative 10; Mismatches 42; Indels 27; Gaps 7;

QY 40 NQFIWKPGWCRGCTGDCPGGVQGRVAVCFHVDGWTSHLSN-----CGESNRPKK 89
Dd 907 SQPVVTEEWGACRSCKGLGVQRTGTCQL-----LPLSNTHKVPKAKA-GDRPEA 959
QY 90 ERSCEPV-CDHNSDLFQVEVSDHHCVLVYPARGE-VKPTAEVCVTAQHGLQH 140
Dd 960 RRPCLRVPCP-----AQWRLGAWSC---SATCGEGIQQRQVVCRTNANSLGH 1004

```

RESULT 10

AT15\_HUMAN

ID AT15\_HUMAN STANDARD; PRT; 950 AA.

AC Q9TES6;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE ADAMTS-15 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 15) (ADAM-TS15) (ADAM-TS15).

GN ADAMTS15

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-21856482; PubMed-11867212;

RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V., Lopez-Otin C.;

RT "Cloning, expression analysis, and structural characterization of seven novel human ADAMTSs, a family of metalloproteinases with disintegrin and thrombospondin-1 domains.";

RT Gene 283:49-62(2002)

RL

CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

CC -!- TISSUE SPECIFICITY: Expressed in fetal liver and kidney, but not in any of the adult tissues examined.

CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important for a tight interaction with the extracellular matrix (By similarity).

CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By similarity).

CC -!- SIMILARITY: Belongs to peptidase family M12B.

CC -!- SIMILARITY: Contains 1 disintegrin-like domain.

CC -!- SIMILARITY: Contains 3 TSP type-1 domains.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL; AJ315733; CAC86014.1; -.

CC MEROPS; M12.025; -.

CC Genew; HGNC:16305; ADAMTS15.

CC MIM; 607509; -.

CC InterPro; IPR001762; Disintegrin.

CC InterPro; IPR001818; Pept\_M10A\_M12B.

CC InterPro; IPR006025; Pept\_M\_Zn\_BS.

CC InterPro; IPR001590; Peptidase\_M12B.

CC InterPro; IPR002870; Peptidase\_M12B\_N.

CC InterPro; IPR000884; TSP1.

CC Pfam; PF01562; Pep\_M12B\_propep; 1.

CC Pfam; PF01421; Reprolysin; 1.

CC Pfam; PF00090; tsp\_1; 3.

CC SMART; SM00209; TSP1; 3.

CC PROSITE; PS02015; ADAM\_MEPRO; 1.

CC PROSITE; PS00546; CYSTEINE\_SWITCH; FALSE\_NEG.

CC PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.

CC PROSITE; PS02014; DISINTEGRIN\_2; FALSE\_NEG.

CC PROSITE; PS00092; TSP1; 3.

CC PROSITE; PS00142; ZINC\_PROTEASE; 1.

CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

CC Repeat; Extracellular matrix.

FT SIGNAL 1 17 POTENTIAL.

FT PROPEP 18 212 BY SIMILARITY.

FT CHAIN 213 950 ADAMTS-15.

FT DOMAIN 213 427 METALLOPROTEASE.

FT DOMAIN 428 515 DISINTEGRIN-LIKE.

FT DOMAIN 516 571 TSP TYPE-1 1.

```

FT DOMAIN 572 700 CYS-RICH.
FT DOMAIN 701 838 SPACER.
FT DOMAIN 839 895 TSP TYPE-1 2.
FT DOMAIN 896 949 TSP TYPE-1 3.
FT SITE 174 174 CYSTEINE SWITCH (POTENTIAL).
FT METAL 361 361 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 362 362 BY SIMILARITY.
FT METAL 365 365 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 950 AA; 103286 MW; 5DFBE18285CCCC3B CRC64;

Query Match 13.4%; Score 111; DB 1; Length 950;
Best Local Similarity 33.0%; Pred. No. 0.0013;
Matches 31; Conservative 13; Mismatches 38; Indels 12; Gaps 6;

QY 44 WKPGPWGCTGDCGPGVQSRVAVCFHVDGWTSHLNCGESNRPKERSCFRVCMDHSDL 103
DB 843 WVAGSWGPGASCG-SGLQKRAVDCRGSAGQRT-VFACDAHRPVETQACGFCP----- 895

QY 104 FQEVSDWHHCVLPVYARGEVKPRACVTAQH 137
DB 896 -TWELSAWSPC-SKSCRG-FQRLSKCV--GHG 924

RESULT 11
ID AT19 HUMAN STANDARD; PRT; 1207 AA.
AC Q8TE59;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAMTS-19 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 19) (ADAM-TS 19) (ADAM-TS19).
DE ADAMTS19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
CC -! COPACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -! SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -! TISSUE SPECIFICITY: Expressed in fetal lung, but not in any adult
CC tissues examined. Expression was detected in an osteosarcoma CDNA
CC library.
CC -! PFM: The precursor is cleaved by a furin endopeptidase (By
CC similarity).
CC -! SIMILARITY: Belongs to peptidase family M12B.
CC -! SIMILARITY: Contains 1 disintegrin-like domain.
CC -! SIMILARITY: Contains 1 PLAC domain.
CC -! SIMILARITY: Contains 5 TSP type-1 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ311904; CAC84565.1; -.

```

```

DR Genew; HGNC:17111; ADAMTS19.
DR MIM; 607513; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001818; Pept_M10A_M12B.
DR InterPro; IPR006925; Pept_M_Zn_BS.
DR InterPro; IPR001590; Peptidase_M12B.
DR InterPro; IPR002870; Peptidase_M12B_N.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF01562; Pcp_M12B_propep; 1.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF00090; Tsp1; 5.
DR SMART; SMO0208; TSP1; 5.
DR PROSITE; PS0215; ADAM_MBP; 1.
DR PROSITE; PS00546; CYSTEINE SWITCH; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS0214; DISINTEGRIN_2; FALSE_NEG.
DR PROSITE; PS00900; PLAC; 1.
DR PROSITE; PS50092; TSP1; 5.
DR PROSITE; PS50092; TSP1; 5.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 316 BY SIMILARITY.
FT CHAIN 317 1207 ADAMTS-19.
FT DOMAIN 317 545 METALLOPROTEASE.
FT DOMAIN 546 633 DISINTEGRIN-LIKE.
FT DOMAIN 634 686 TSP TYPE-1 1.
FT DOMAIN 687 790 CYS-RICH.
FT DOMAIN 791 914 SPACER.
FT DOMAIN 915 975 TSP TYPE-1 2.
FT DOMAIN 976 1037 TSP TYPE-1 3.
FT DOMAIN 1039 1083 TSP TYPE-1 4.
FT DOMAIN 1087 1144 TSP TYPE-1 5.
FT DOMAIN 1160 1199 PLAC.
FT DOMAIN 88 216 PRO-RICH.
FT SITE 294 294 CYSTEINE SWITCH (POTENTIAL).
FT METAL 482 482 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 483 483 BY SIMILARITY.
FT METAL 486 486 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 492 492 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 797 797 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 907 907 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1009 1009 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1207 AA; 134061 MW; AF36F6BF5886FDE2 CRC64;

Query Match 13.3%; Score 110.5; DB 1; Length 1207;
Best Local Similarity 27.3%; Pred. No. 0.0018;
Matches 30; Conservative 19; Mismatches 44; Indels 17; Gaps 7;

QY 31 AAHLGKKNQPIWKPFGWGRCTGDCGPGVQSRVAVCFHVDGWTSHLNCGESNRPKPE 90
DB 1028 AQRCQG-QDCMTVWEAGVWSFYSVKCGK-GIRHRTVC-----TNPRKCVLSTRPRA 1079

QY 91 RSCFRVCMDHSDLPQWESVDHHCVLVYARGEVKPRAC---VTAQH 137
DB 1080 ED---CEDYSKYVWRMGDWSKC-SITCGKG-MQSRVIOCMHKITGRHG 1123

RESULT 12
ATSL_RAT
ID ATSL_RAT STANDARD; PRT; 967 AA.
AC Q9WUQ1; Q9ERI1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
DE ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```



NCBI\_TaxID=10116;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN-Sprague-Dawley; TISSUE-Brain;  
 Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,  
 Little S.P.;  
 "Induction of a disintegrin and metalloprotease with the  
 thrombospondin type I motif (ADAMTS).";  
 Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 [2]  
 SEQUENCE OF 18-967 FROM N.A.  
 STRAIN-Sprague-Dawley; TISSUE-Liver;  
 MEDLINE=20304099; PubMed=10847486;  
 Diamantis I., Luethi M., Hoelsi M., Reichen J.;  
 "Cloning of the rat ADAMTS-1 gene and its down regulation in  
 endothelial cells in cirrhotic rats.";  
 Liver 20:165-172(2000).  
 CC -!- FUNCTION: Cleaves aggrecan, a cartilage proteoglycan, and may be  
 involved in its turnover. Has angiogenic inhibitor activity (By  
 similarity). Active metalloprotease, which may be associated with  
 various inflammatory processes as well as development of cancer  
 cachexia. May play a critical role in follicular rupture (By  
 similarity).  
 CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 1683-Glu-|-Leu-1684  
 site, within the chondroitin sulfate attachment domain.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
 matrix (By similarity).  
 CC -!- INDUCTION: Down-regulated in endothelial cells derived from  
 cirrhotic liver.  
 CC -!- DOMAIN: The spacer domain and the TSP type-1 domains are important  
 for a tight interaction with the extracellular matrix.  
 CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By  
 similarity).  
 CC -!- SIMILARITY: Belongs to peptidase family M12B.  
 CC -!- SIMILARITY: Contains 1 disintegrin-like domain.  
 CC -!- SIMILARITY: Contains 3 TSP type-1 domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC  
 CC EMBL; AF149118; RAD34012.1; -;  
 CC EMBL; AF304446; ARG29823.1; -;  
 CC MEROPS; M12.222; -;  
 CC InterPro; IPR006586; ADAM\_cysteine.  
 CC InterPro; IPR001762; Disintegrin.  
 CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro; IPR001590; Peptidase\_M12B.  
 CC InterPro; IPR002870; Peptidase\_M12B\_N.  
 CC InterPro; IPR000884; TSP1.  
 CC InterPro; IPR008085; TSP 1.  
 CC Pfam; PF01562; Pep\_M12B\_propep; 1.  
 CC Pfam; PF01421; Reprolysin; 1.  
 CC Pfam; PF00090; tsp 1; 3.  
 CC PRINTS; PR01705; TSP1REPEAT.  
 CC SMART; SM00608; ACR; 1.  
 CC SMART; SM00209; TSP1; 3.  
 CC PROSITE; PS50215; ADAM\_MEPRO; 1.  
 CC PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
 CC PROSITE; PS50214; DISINTEGRIN\_2; FALSE\_NEG.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC PROSITE; PS50092; TSP1\_3.  
 CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix; Heparin-binding.  
 KW SIGNAL 1 54  
 FT PROPEP 55 252 POTENTIALITY.  
 FT CHAIN 253 967 ADAMTS-1.  
 FT DOMAIN 253 475 METALLOPROTEASE.

FT DOMAIN 476 558 DISINTEGRIN-LIKE.  
 FT TSP TYPE-1 1.  
 FT DOMAIN 559 614  
 FT DOMAIN 616 724  
 FT CYS-RICH.  
 FT DOMAIN 725 857  
 FT SPACER.  
 FT DOMAIN 854 910  
 FT TSP TYPE-1 2.  
 FT DOMAIN 911 967  
 FT TSP TYPE-1 3.  
 FT DOMAIN 194 198  
 FT POLY-ARG.  
 FT SITE 205 205 CYSTEINE SWITCH (POTENTIAL).  
 FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT SITE 402 402 BY SIMILARITY.  
 FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 21 21 I -> V (IN REF. 2).  
 FT CONFLICT 26 31 KERSQ -> RSRGSL (IN REF. 2).  
 FT CONFLICT 49 49 V -> A (IN REF. 2).  
 FT CONFLICT 72 72 R -> P (IN REF. 2).  
 FT CONFLICT 79 79 L -> TR (IN REF. 2).  
 FT CONFLICT 249 249 R -> G (IN REF. 2).  
 FT CONFLICT 262 265 TMLV -> NLK (IN REF. 2).  
 FT CONFLICT 607 607 S -> F (IN REF. 2).  
 FT CONFLICT 936 936 L -> V (IN REF. 2).  
 FT CONFLICT 962 962 I -> T (IN REF. 2).  
 SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCB4CF CRC64;  
 Query Match 13.3%; Score 110; DB 1; Length 967;  
 Best Local Similarity 27.9%; Pred. No. 0.0016;  
 Matches 36; Conservative 17; Mismatches 48; Indels 28; Gaps 8;  
 QY 26 LLSHA-----AHLEKKNQF-----TWKPGWRCCTGDCGPGVQSRVWCFHV 71  
 DB 826 LMVGHALRPKIKYTFPMKKKTFEFAIPTFSEWIEWGECKSTCG-SGQRRVVECRDI 884  
 QY 72 DGMTSHLSNGESNRPPKERSFRVCDWHSDF--PQWEVSDWHHCVLVYARGEVKPRTA 129  
 DB 885 NGHPA--SECAKEVKPASTRPC-----ADLPGRWQVGDWSPCSKT-CGKG-YKCRTL 933  
 QY 130 ECVTAQHGL 138  
 DB 934 KCLSHDGGV 942  
 RESULT 13  
 ATSP\_HUMAN  
 ID ATSP\_HUMAN STANDARD; PRT; 1211 AA.  
 AC O95450;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase  
 with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2) (Procollagen I/II  
 amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC  
 I-NP) (Procollagen N-endopeptidase) (pNPI).  
 GN ADAMTS2 OR PCINP OR PCPNI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS LNP1 AND SPNPI), AND, DISEASE.  
 RC TISSUE=Skin;  
 RX MEDLINE=99347935; PubMed=10417273;  
 RA Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,  
 RA Warteleski W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,  
 RA Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.;  
 RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis  
 are caused by mutations in the procollagen I N-proteinase gene.";  
 RL Am. J. Hum. Genet. 65:308-317(1999).  
 CC -!- FUNCTION: Cleaves the propeptides of type I and II collagen prior

to fibril assembly. Does not act on type III collagen. May also play a role in development that is independent of its role in collagen biosynthesis.

-!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain at Ala-I(I)-at Pro-I-Gln and of alpha-1(I) and alpha-2(I) chains at Ala-I-Gln.

-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

-!- SUBUNIT: May belong to a multimeric complex. Binds specifically to collagen type XIV (By similarity).

-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

-!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=2;

Name=LGNPI;

isoId=O95450-1; Sequences=Displayed;

Name=SPNPI;

isoId=O95450-2; Sequences=VSP\_005497, VSP\_005498;

Note=Has no significant N-procollagen peptidase activity;

-!- TISSUE SPECIFICITY: Expressed at high level in skin, bone, tendon and aorta and at low levels in thymus and brain.

-!- DOMAIN: The spacer domain and the TSP type-1 domains are important for a tight interaction with the extracellular matrix.

-!- PTM: The precursor is cleaved by a furin endopeptidase (By similarity).

-!- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos syndrome type VIIC (EDS VIIC) [MIM:225410]. EDS VIIC is a recessively inherited connective-tissue disorder characterized clinically by severe skin fragility and joint hypermobility and biochemically by the presence in skin of procollagen incompletely processed at the N-terminus.

-!- SIMILARITY: Belongs to peptidase family M12B.

-!- SIMILARITY: Contains 1 disintegrin-like domain.

-!- SIMILARITY: Contains 4 TSP type-1 domains.

-!- CAUTION: Has sometimes been referred to as ADAMTS3.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL: AJ003125; CAA05880.1; -

DR MEROPS; M12.301; -

DR Genew; HGNC:218; ADAMTS2.

DR MIM; 604539; -

DR MIM; 225410; -

DR GO; GO:0008133; F:collagenase activity; TAS.

DR GO; GO:0008237; F:metallopeptidase activity; TAS.

DR InterPro; IPR001762; Disintegrin.

DR InterPro; IPR006025; Rept\_Mn\_BS.

DR InterPro; IPR001590; Peptidase\_M12B.

DR InterPro; IPR002870; Peptidase\_M12B\_N.

DR InterPro; IPR000884; TSP1.

DR InterPro; IPR008085; TSP1.

DR Pfam; PF01562; Pep\_M12B\_propep; 1.

DR Pfam; PF01421; Repolysin; 1.

DR Pfam; PF00090; tsp\_1; 4.

DR PRINTS; PR01705; TSP1REPEAT.

DR SMART; SM00209; TSP1; 4.

DR PROSITE; PS02015; ADAM\_MPRO; 1.

DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE NEG.

DR PROSITE; PS02014; DISINTEGRIN\_2; FALSE\_NEG.

DR PROSITE; PS00900; PLAC; 1.

DR PROSITE; PS00092; TSP1; 4.

DR PROSITE; PS00142; ZINC\_PROTEASE; FALSE NEG.

DR PROSITE; PS00142; ZINC\_PROTEASE; FALSE NEG.

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat; Collagen degradation; Extracellular matrix;

KW Alternative splicing; Ehlers-Danlos syndrome.

FT SIGNAL 1 29 POTENTIAL.

FT PROPEP 30 253 BY SIMILARITY.

CHAIN	254	1211	ADAMTS-2.
FT DOMAIN	254	479	METALLOPROTEASE.
FT DOMAIN	560	560	DISINTEGRIN-LIKE.
FT DOMAIN	561	616	TSP TYPE-1 1.
FT DOMAIN	618	722	CYS-RICH.
FT SITE	691	693	CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN	723	851	SPACER.
FT DOMAIN	854	912	TSP TYPE-1 2.
FT DOMAIN	914	971	TSP TYPE-1 3.
FT DOMAIN	975	1029	TSP TYPE-1 4.
FT DOMAIN	1059	1097	PLAC.
FT DOMAIN	40	43	POLY-ALA.
FT METAL	185	188	POLY-GLU.
FT ACT SITE	409	409	ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL	412	412	ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL	418	418	ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD	112	112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	251	251	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	949	949	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	993	993	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	1031	1031	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	1098	1098	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	1145	1145	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	1150	1150	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC	544	566	HCPKRGHCWLPDLKKDGSWGA -> FRGAVAHACYPST
FT			LGGQGRWIA (in isoform SpNP1).
FT			/FtId=VSP_005497.
FT			Missing (in isoform SpNP1).
FT			/FtId=VSP_005498.
SQ	SEQUENCE	1211 AA; 134722 MW; BECEE25C23CAD2D CRC64;	
Query Match	13.2%;	Score 109.5; DB 1; Length 1211;	
Best Local Similarity	28.0%;	Pred. No. 0.0023;	
Matches	30;	Conservative 12; Mismatches 38; Indels 27; Gaps 6;	
QY	40	NQFTWKCPGRCRTGDCPGVQSRVAVWCFHV--DGWTS--HLSNCGESNPPKRCSCFR 95	
Db	914	SQPVVVTGEMPECSQTCGTGMQVRSVRCIPLHNTTRSHVAKHCNDA-RPESRRACSR 972	
QY	96	VCWHMSDLFWQVSDWHHCVLVPYARGEVKPRTAEVCVTAHQGLQHR 141	
Db	973	ELCPG-----RWAGPWSQC-----SVTCGNGTQER 998	
RESULT 14			
AT10_HUMAN			
ID	AT10_HUMAN	STANDARD;	PRT; 1077 AA.
AC	Q9H324;		
DT	28-FEB-2003	(Rel. 41, Created)	
DT	28-FEB-2003	(Rel. 41, Last sequence update)	
DT	15-MAR-2004	(Rel. 43, Last annotation update)	
DE	ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).		
GN	ADAMTS10.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]_TaxID=9606;		
RP	SEQUENCE FROM N.A.		
RA	Apte S.S.;		
RT	"ADAM-TS10; a novel member of the ADAM-TS family containing multiple thrombospondin type I repeats."		
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
CC	-!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).		
CC	-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).		
CC	-!- DOMAIN: The spacer domain and the TSP type-1 domains are important for a tight interaction with the extracellular matrix (By similarity).		
CC	-!- SIMILARITY: Belongs to peptidase family M12B.		
CC	-!- SIMILARITY: Contains 1 disintegrin-like domain.		

CC -1- SIMILARITY: Contains 1 PLAC domain.  
CC -1- SIMILARITY: Contains 5 TSP type-1 domains.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF163762; AAC35563.1; -  
CC DR MEROPS; M12.235; -  
CC DR Genew; HGNC:13201; ADAMTS10.  
CC DR InterPro; IPR001762; Disintegrin.  
CC DR InterPro; IPR006025; Pept\_M\_2n\_BS.  
CC DR InterPro; IPR001590; Peptidase\_M12B.  
CC DR InterPro; IPR002870; Peptidase\_M12B\_N.  
CC DR InterPro; IPR000884; TSP1.  
CC DR InterPro; IPR008085; TSP\_1.  
CC DR Pfam; PF01562; Pep\_M12B\_propep; 1.  
CC DR Pfam; PF01421; Reptolysin; 1.  
CC DR Pfam; PF00090; TSP\_1; 5.  
CC DR PRINTS; PR01705; TSP1REPEAT.  
CC DR SMART; SM00209; TSP1; 5.  
CC DR PROSITE; PS00215; ADAM\_MEPRO; 1.  
CC DR PROSITE; PS00427; DISINTEGRIN\_1; FALSE\_NEG.  
CC DR PROSITE; PS0214; DISINTEGRIN\_2; FALSE\_NEG.  
CC DR PROSITE; PS05900; PLAC; 1.  
CC DR PROSITE; PS50092; TSP1; 5.  
CC DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
CC DR Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;  
CC Repeat; Extracellular matrix.  
CC FT NON TER 1  
CC FT PROPEP <1 207 BY SIMILARITY.  
CC FT CHAIN 208 1077 ADAMTS-10.  
CC FT DOMAIN 208 433 METALLOPROTEASE.  
CC FT DOMAIN 434 520 DISINTEGRIN-LIKE.  
CC FT DOMAIN 578 679 CIS-RICH.  
CC FT DOMAIN 680 802 SPACER.  
CC FT DOMAIN 521 576 TSP TYPE-1 1.  
CC FT DOMAIN 799 857 TSP TYPE-1 2.  
CC FT DOMAIN 858 919 TSP TYPE-1 3.  
CC FT DOMAIN 921 975 TSP TYPE-1 4.  
CC FT DOMAIN 977 1032 TSP TYPE-1 5.  
CC FT DOMAIN 1039 1077 PLAC.  
CC FT METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).  
CC FT ACT SITE 367 367 ZINC (CATALYTIC) (BY SIMILARITY).  
CC FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).  
CC FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).  
CC FT CARBOHYD 54 64 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 714 714 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 769 769 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCBFF587 CRC64;  
Query Match 13.1%; Score 108.5; DB 1; Length 1077;  
Best Local Similarity 41.8%; Pred. No. 0.0026;  
Matches 23; Conservative 4; Mismatches 25; Indels 3; Gaps 2;  
QY 44 WKPGWGRGTCGGPGVQSGVAVWCFHVDGWTSHLNCGESNRPPKSCRFVCD 98  
DB 981 WVAGEGECACGGVQGRQ-RSVRCTSHGTQASH--ECTEALRPPTTQCCCAKCD 1032  
RESULT 15  
ATS2 BOVIN STANDARD; PRT; 1205 AA.  
AC P79331;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase  
DE with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2) (Procollagen I/II  
DE amino-propeptide processing enzyme) (Procollagen I N-proteinase) (PC  
DE I-NP) (Procollagen N-endopeptidase) (pNPI).  
GN ADAMTS2 OR NPI.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OC NCBI\_TaxID=9913;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Skin;  
RX MEDLINE=97225960; PubMed=9122202;  
RA Collige A., Li S.W., Sieron A.L., Nuegens B.V., Prockop D.J.,  
RA Lapiere C.M.;  
RT "cDNA cloning and expression of bovine procollagen I N-proteinase: a  
RT new member of the superfamily of zinc-metalloproteinases with binding  
RT sites for cells and other matrix components.";  
RL Prog. Natl. Acad. Sci. U.S.A. 94:2374-2379 (1997).  
RN [2]  
RN PARTIAL SEQUENCE.  
RP MEDLINE=95348096; PubMed=7622483;  
RA Collige A., Beschin A., Samyn B., Goebels Y., Van Beeumen J.,  
RA Nuegens B.V., Lapiere C.M.;  
RT "Characterization and partial amino acid sequencing of a 107-kDa  
RT procollagen I N-proteinase purified by affinity chromatography on  
RT immobilized type XIV collagen.";  
RL J. Biol. Chem. 270:16724-16730 (1995).  
CC -1- FUNCTION: Cleaves the propeptides of type I and II collagen prior  
CC to fibril assembly. Does not act on type III collagen. May also  
CC play a role in development that is independent of its role in  
CC collagen biosynthesis.  
CC -1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain  
CC at alpha-1(I) at Pro-[Gln and of alpha-1(II) and alpha-2(I) chains  
CC at Ala-[Gln.  
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -1- SUBUNIT: May belong to a multimeric complex. Binds specifically to  
CC collagen type XIV.  
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular  
CC matrix (By similarity).  
CC -1- TISSUE SPECIFICITY: Enzymatic activity is detected at high level  
CC in all type I collagen-rich tissues such as skin, bones, tendons  
CC and aorta and at low level in brain and thymus. The mRNA levels  
CC were disproportionately high in heart, liver, retina and muscle.  
CC -1- DOMAIN: The spacer domain and the TSP type-1 domains are important  
CC for a tight interaction with the extracellular matrix.  
CC -1- PTM: The N-terminus is blocked.  
CC -1- PTM: The precursor is cleaved by a furin endopeptidase (By  
CC similarity).  
CC -1- DISEASE: Defects in ADAMTS2 are the cause of dermatosparaxis, a  
CC recessively inherited disorder characterized by severe skin  
CC fragility and biochemically by the presence in skin of procollagen  
CC incompletely processed at the N-terminus.  
CC -1- SIMILARITY: Belongs to peptidase family M12B.  
CC -1- SIMILARITY: Contains 1 disintegrin-like domain.  
CC -1- SIMILARITY: Contains 1 PLAC domain.  
CC -1- SIMILARITY: Contains 4 TSP type-1 domains.  
CC -1- CAUTION: Has sometimes been referred to as ADAMTS3.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X96389; CAA65253.1; -  
CC DR PIR; T18517; T18517.  
CC DR MEROPS; M12.301; -  
CC DR InterPro; IPR001762; Disintegrin.

```
OR InterPro; IPR006025; Pept_M_Zn_BS.
OR InterPro; IPR001590; Peptidase_M12B.
OR InterPro; IPR002870; Peptidase_M12B_N.
OR InterPro; IPR000884; TSP1.
OR Pfam; PF01562; Pept_M12B_propep; 1.
OR Pfam; PF01421; Reprolysin; 1.
OR Pfam; PF00090; tsp_1; 4.
OR SMART; SM00209; TSP1; 4.
OR PROSITE; PS0215; ADAM_MEPRO; 1.
OR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
OR PROSITE; PS0214; DISINTEGRIN_2; FALSE_NEG.
OR PROSITE; PS0300; PLAC; 1.
OR PROSITE; PS0092; TSP1; 4.
OR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
OR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
OR Repeat; Collagen degradation; Extracellular matrix.
OR SIGNAL 1 28 POTENTIAL.
OR PROPEP 29 253 BY SIMILARITY.
OR CHAIN 254 1205 ADAMS-2.
OR DOMAIN 254 473 METALLOPROTEASE.
OR DOMAIN 474 554 DISINTEGRIN-LIKE.
OR DOMAIN 555 610 TSP TYPE-1 1.
OR DOMAIN 612 716 CYS-RICH.
OR DOMAIN 717 845 SPACER.
OR DOMAIN 848 906 TSP TYPE-1 2.
OR DOMAIN 908 968 TSP TYPE-1 3.
OR DOMAIN 969 1023 TSP TYPE-1 4.
OR DOMAIN 1053 1091 PLAC.
OR DOMAIN 31 35 POLY-ALA.
OR DOMAIN 177 180 POLY-GLU.
OR METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).
OR ACT_SITE 403 403 BY SIMILARITY.
OR METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
OR METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
OR SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).
OR CARBOHYD 104 104 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 942 942 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 943 943 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 987 987 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 1025 1025 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 1092 1092 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 1139 1139 N-LINKED (GLCNAC. .) (POTENTIAL).
OR CARBOHYD 1144 1144 N-LINKED (GLCNAC. .) (POTENTIAL).
OR SEQUENCE 1205 AA; 133887 MW; 735B232A45320371 CRC64;
Query Match 13.0%; Score 107.5; DB 1; Length 1205;
Best Local Similarity 28.0%; Pred. No. 0.0037;
Matches 30; Conservative 13; Mismatches 37; Indels 27; Gaps 6;
2Y 40 NQFIWKPGMGRCGTGCGGGVQSVAVWC---PHVDGWTSLHLSNCGESNRPPKERSQPR 95
DB 908 SQPVWVTGWEPCRSRGRTGMQRVRCVQPLHNTTTSVHTKCNDA-RPEGRACNR 966
2Y 96 -VCDWHSDLFQWEVSDWHHCVLVPYARGEYKPTACVTAQHGLOHR 141
DB 967 ELCPG-----RWAGWSQC-----SVTCNGTQER 992
```

Search completed: February 24, 2004, 01:03:08  
Job time : 4.97635 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: February 24, 2004, 00:30:51 ; Search time 10.692 Seconds  
(without alignments)  
4190.395 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MFPSNLTVCWWSRWK.....EVKPTACVTAQHGLQHRM 142

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_podent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	667	80.4	1107	11 Q8BHP3	Q8bhp3 mus musculus
2	414	49.9	1536	4 Q9C014	Q9c014 homo sapien
3	127	15.3	622	4 Q9H8X0	Q9h8x0 homo sapien
4	126	15.2	1280	11 Q9EPX2	Q9epx2 mus musculus
5	125	15.1	701	11 Q8CDB8	Q8cdb8 mus musculus
6	124	14.9	711	5 Q8I411	Q8i411 caenorhabdi
7	121.5	14.6	957	11 Q7TSK7	Q7tsk7 mus musculus
8	120.5	14.5	1235	4 Q95428	Q95428 homo sapien
9	119.5	14.4	951	4 Q60345	Q60345 homo sapien
10	119.5	14.4	951	4 Q86TH1	Q86th1 homo sapien
11	118	14.2	724	5 Q15916	Q15916 dictyostell
12	117.5	14.2	898	4 Q9UFZ4	Q9ufz4 homo sapien
13	116.5	14.0	340	11 Q91Z56	Q91z56 mus musculus
14	116	14.0	237	4 Q9HBS6	Q9hbs6 homo sapien
15	116	14.0	424	4 Q8N643	Q8n643 homo sapien
16	114.5	13.8	640	11 Q8BSM2	Q8bsm2 mus musculus

17	114.5	13.8	1043	5 P30884	P30884 caenorhabdi
18	112	13.5	235	11 Q91YP4	Q91yp4 mus musculus
19	112	13.5	1036	11 Q80T21	Q80t21 mus musculus
20	111.5	13.4	1502	4 Q9UP26	Q9up26 homo sapien
21	111.5	13.3	1092	11 Q8BXA1	Q8bxa1 mus musculus
22	110.5	13.2	2165	5 Q19751	Q19751 caenorhabdi
23	109	13.1	1023	4 Q9ULI7	Q9uli7 homo sapien
24	107	12.9	759	6 Q8H2M8	Q8h2m8 equus cabal
25	106.5	12.8	446	4 Q43384	Q43384 homo sapien
26	106.5	12.8	740	4 Q8WUV0	Q8wuv0 homo sapien
27	105.5	12.7	310	11 Q8C9R1	Q8c9r1 mus musculus
28	105	12.7	3198	5 Q9U8G8	Q9u8g8 manduca sex
29	103.5	12.5	432	4 Q9NPM2	Q9npm2 homo sapien
30	103.5	12.5	2174	5 Q9GQK0	Q9gqk0 drosophila
31	103.5	12.5	2772	5 Q9AV4	Q9av4 drosophila
32	103.5	12.5	2776	5 Q869A0	Q869a0 drosophila
33	103.5	12.5	2898	5 Q868Z9	Q868z9 drosophila
34	103	12.4	417	4 Q9BXV3	Q9bxv3 homo sapien
35	103	12.4	525	4 Q96RW4	Q96rw4 homo sapien
36	103	12.4	1487	5 Q8MPV5	Q8mpv5 caenorhabdi
37	103	12.4	1558	5 Q8I710	Q8i710 caenorhabdi
38	103	12.4	2167	5 Q76840	Q76840 caenorhabdi
39	102.5	12.3	364	4 Q96AY5	Q96ay5 homo sapien
40	101	12.2	538	4 Q9H8E4	Q9h8e4 homo sapien
41	100	12.0	1009	11 Q8BKY1	Q8bky1 mus musculus
42	100	12.0	1600	11 Q811B3	Q811b3 mus musculus
43	99.5	12.0	377	11 Q8BLI0	Q8bli0 mus musculus
44	99.5	12.0	761	6 Q95LQ2	Q95lq2 macaca fasc
45	98.5	11.9	192	11 Q9CX59	Q9cx59 mus musculus

#### ALIGNMENTS

RESULT 1

Q8BHP3 PRELIMINARY; PRT; 1107 AA.

AC Q8BHP3;  
DT 01-MAR-2003 (TRENBLrel. 23, Created)  
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Weakly similar to hypothetical protein DJ110N13.1 in chromosome 7.  
GN D130067103RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL; AK051714; BAC34732.1; -  
DR MGD; MGI:2443925; D130067103RIK.  
DR GO; GO:0016030; C:membrane; IEA.  
DR GO; GO:0004691; F:oxidoreductase activity; IEA.  
DR GO; GO:0006618; P:electron transport; IEA.  
DR InterPro; IPR005797; Cytb\_b6\_N.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; tsp\_1; 7.  
DR SMART; SM00209; TSP1; 9.  
DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
DR PROSITE; PS00092; TSP1; 10.  
KW Hypothetical protein.  
SQ SEQUENCE 1107 AA; 123640 MW; B34C708E6AA65734 CRC64;

Query Match 80.4%; Score 667; DB 11; Length 1107;  
Best Local Similarity 80.1%; Pred. No. 2.7e-67;

```

Matches 113; Conservative 13; Mismatches 15; Indels 0; Gaps 0;
QY 1 MFPSNLTVTCWVRSWEKFLLLSLLSHAHLEGGKDNQFIWKPGRCTGDCGPGG 60
DB 1 MFLKSDLAVTWVRSWEKFLVLSLLSQAHLLEGRDNQFLWKTGPWRCAGDCGPGG 60
QY 61 VQSRVAVMCFHVDGWTSHLSNCGESNRPPKERSCFRVCDDHSDLFQWEVSDWHHCVLVPPYA 120
DB 61 AQSRVAVMCFHIEGWTSPMSNCDSESSQPPKERSCFRVCDDHSDLFQWEVSDWHHCVLVPPYA 120
QY 121 RGEVPRTAECVTAQHGLQHR 141
DB 121 QGEPRPRAVECVTAQHGLQHR 141

RESULT 2
QSC014
ID Q9C014 PRELIMINARY; PRT; 1536 AA.
AC Q9C014;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein KIAA1679 (Fragment).
GN KIAA1679.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
DR EMBL; AB051466; BAB21770.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 8.
DR SMART; SM00209; TSP1; 16.
DR PROSITE; PS50092; TSP1; 12.
KW Hypothetical protein.
FT NON TER 1
RN SEQUENCE 1536 AA; 171242 MW; B5121C8160EEDCE CRC64;

Query Match 49.9%; Score 414; DB 4; Length 1536;
Best Local Similarity 100.0%; Pred. No. 3e-38;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 73 GWTSHLSNCGESNRPPKERSCFRVCDDHSDLFQWEVSDWHHCVLVPPYARGEVKPRTAECV 132
DB 1 GWTSHLSNCGESNRPPKERSCFRVCDDHSDLFQWEVSDWHHCVLVPPYARGEVKPRTAECV 60
QY 133 TAQHGLQHRM 142
DB 61 TAQHGLQHRM 70

RESULT 3
Q9H8X0
ID Q9H8X0 PRELIMINARY; PRT; 622 AA.
AC Q9H8X0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ13166.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,

```

```

RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NECD human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK023228; BAB14479.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 6.
DR SMART; SM00209; TSP1; 8.
DR PROSITE; PS50092; TSP1; 7.
KW Hypothetical protein.
SQ SEQUENCE 622 AA; 70229 MW; DC415F663E5B8838 CRC64;

Query Match 15.3%; Score 127; DB 4; Length 622;
Best Local Similarity 29.7%; Pred. No. 7.1e-06;
Matches 30; Conservative 15; Mismatches 38; Indels 18; Gaps 5;
QY 39 DNQFI-----WKPGPWGRCTGCGPGVQSRVAVMCFHVDGWTSHLSNCGESNRPPKE 90
DB 413 ENQVHPVSRGNGWRTGPGWGCSSCS--GGLQRAVVCQDENGQSA--SYCDAASKPPPEL 469
QY 91 RSCFRVCDHSDLFQWEVSDWHHCVLVPPYARGEVKPRTAEC 131
DB 470 QQC-----GPGCPQWNYGNWGCESQT--CGGKISRLVIC 503

RESULT 4
Q9EPX2
ID Q9EPX2 PRELIMINARY; PRT; 1280 AA.
AC Q9EPX2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Papilin.
GN PAPLN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1.
RX MEDLINE=20530499; PubMed=11076767;
RA Kramarova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Kramarov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,
RA Sieron A.L., Prockop D.J., Fessler J.H.;
RT "Papilin in development; a pericellular protein with a homology to the
RT ADAMTS metalloproteinases.";
RL Development 127:5475-5485(2000).
DR EMBL; AF314171; AAG41980.1; -.
DR HSSP; P12111; 2KNT.
DR MGD; MGI:2386139; Papln.
DR GO; GO:0003824; F: catalytic activity; IEA.
DR GO; GO:0004867; F: serine protease inhibitor activity; IEA.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG c2.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; IG; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp.1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS50279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS50835; IG_LIKE; 3.

```

```
DR PROSITE; PS00879; ODR_DC_2_2; 1.
DR PROSITE; PS50092; TSP1; 5.
KW Immunoglobulin domain; Protease inhibitor; Serine protease inhibitor.
SQ SEQUENCE 1280 AA; 138824 MW; A2287705E561AF30 CRC64;

Query Match
Best Local Similarity 15.2%; Score 126; DB 11; Length 1280;
Matches 31; Conservative 15; Mismatches 39; Indels 10; Gaps 6;

QY 42 FIKPGGWRGCTGDCGGVQSVRAVWC-FHVDGWTSHLSNCGSNPPPKRSC-FRVCWM 99
Db 307 FSHSGSWGDCSRECG-GGHSRLVFTCTINEAYPDHM--CQHPPTHRSCNTQPC-- 361

QY 100 HSDLFQWESVDMHCVLPVARGVKPRTAECVTA 134
Db 362 -PKTRKWKVGFWPTPCSV--SCGGGVQSVRSVYCISS 393

RESULT 5
Q8CDB8 PRELIMINARY; PRT; 701 AA.
AC Q8CDB8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN 8430403M15RIK OR AW743315.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA THE FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
DR EMBL; AK030699; BAC27087.1; -.
DR MGD; MGI:2141719; 8430403M15RIK.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 8.
DR SMART; SM00209; TSP1; 8.
DR PROSITE; PS50092; TSP1; 8.
KW Hypothetical protein.
SQ SEQUENCE 701 AA; 77924 MW; C854E56D0704805F CRC64;

Query Match
Best Local Similarity 15.1%; Score 125; DB 11; Length 701;
Matches 29; Conservative 14; Mismatches 37; Indels 12; Gaps 5;

QY 44 WKPGWRCRTGDCGGVQSVRAVWC-FHVDGWTSHLSNCGSNPPPKRSC-FRVCWHSD 102
Db 125 WRTGFWGACSTCA-GGQRRVVVDCDNGYTA--NDCVERIKPDEQRACSGFCP---- 177

QY 103 LFQWESVDMHCVLPVARGVKPRTAECVTA 134
Db 178 --QWYSGWGECKL--CGGMRTLVVVCRA 205

RESULT 6
Q8I4I1 PRELIMINARY; PRT; 711 AA.
AC Q8I4I1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein F53B6.2b.
GN F53B6.2 OR F53B6.2b.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
```

```
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; Z81086; CAD56586.1; -.
DR WormPep; F53B6.2b; CE32429.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00047; IG; 1.
DR Pfam; PF00090; tsp.1; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00408; IGC2; 1.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS50092; TSP1; 5.
KW Hypothetical protein.
SQ SEQUENCE 711 AA; 79945 MW; 9695028B9636AE94 CRC64;

Query Match
Best Local Similarity 14.9%; Score 124; DB 5; Length 711;
Matches 38; Conservative 11; Mismatches 48; Indels 32; Gaps 6;

QY 17 MKXJFLLLSLLSHAHLE-----GKXDNQFIWK-----PGWRCRTGDCGPG 59
Db 1 MLPLLLSAPLVGSFALMDNDRAYLDGNTFGSDMNDNDWQAPRLVAGWSTCSCTCT- 59

QY 60 GVQSVRAVWC-FHVDGWTSC-----HLSNCGSNRPPKRSK-FRVCWHSDL-----FQ 105
Db 60 GWSRTVECVAVNPISAPIKLPMSQCQEQPKLFSCEVRSCLQEDSKLSDEAPYQ 119

QY 106 WEVSDWHRC 114
Db 120 WRYGDWTCQ 128

RESULT 7
Q7TSK7 PRELIMINARY; PRT; 957 AA.
AC Q7TSK7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TSP1-repeats containing protein.
GN TCP-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Spleen;
RA Kishi Y., Toji S., Tanaka M., Miyajima A., Yahara I.;
RT "Cloning of TSP1-repeats-containing protein expressed in splenic CD11c+
RT cells.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB096928; BAC76875.1; -.
SQ SEQUENCE 957 AA; 105646 MW; D92FB05E0CF7FEC0 CRC64;

Query Match
Best Local Similarity 14.6%; Score 121.5; DB 11; Length 957;
Matches 29; Conservative 10; Mismatches 38; Indels 11; Gaps 4;
```

```
QY 31 AAHLEGGKDNQFIWKPQGWGRTGDCGGVQSVRAVWCFHVDGWTSHLSN---CGESNRP 87
DB 791 AIHPGCDKNCFAHWLAQDWERCNTTCG-RGVKKLVLCMELANGKQIRSGPEGGLAKRP 849
QY 88 PKERSCF-RVCDWHSDDLQWEVSDWHHC 114
DB 850 PEESTCFERPC-----FKWYTPWSEC 871

RESULT 8
O95428 PRELIMINARY; PRT; 1235 AA.
AC O95428;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Dors M., Ratcliffe A.,
RA Madan A., Dickhoff R., Shafer T., James R., Lasky S., Hood L.;
RT "Complete sequence of the gene for presenilin 1.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109907; AAC97963.1; -.
DR HSSP; P12111; 2KNT.
DR GO; GO:0004867; F:serine protease inhibitor activity; IEA.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00047; IG; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00408; IGc2; 3.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSPI; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS0835; IG LIKE; 3.
DR PROSITE; PS0092; TSPI; 5.
KW Hypothetical protein; Immunoglobulin domain; Protease inhibitor;
KW Serine protease inhibitor.
SQ SEQUENCE 1235 AA; 133477 MW; A0B44CCE4F38E350 CRC64;

Query Match 14.5%; Score 120.5; DB 4; Length 1235;
Best Local Similarity 37.5%; Pred. No. 8.2e-05;
Matches 27; Conservative 5; Mismatches 37; Indels 3; Gaps 3;

QY 44 WKPQGWGRTGDCGGVQSVRAVWCFHVDGWTSHLSNCGESNRPKERSCF-RVCDWHSDDLQWEVSDWHHC 102
DB 449 WSPFEPWGESVCGV-GVRKRSVTCRGERGSLHTAACSLDRPPLTPFCVHEDCPILLSD 507
QY 103 LFQWEVSDWHHC 114
DB 508 Q-AWHVGTWGLC 518

RESULT 9
O60345 PRELIMINARY; PRT; 951 AA.
AC O60345;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein KIAA0605.
GN KIAA0605.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628591;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
DR EMBL; AB011177; BAA25531.1; -.
DR FIR; T0260; T00260.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; tsp_1; 5.
DR SMART; SM00209; TSPI; 7.
DR PROSITE; PS0092; TSPI; 4.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 104678 MW; CC73753F232BC1A2 CRC64;

Query Match 14.4%; Score 119.5; DB 4; Length 951;
Best Local Similarity 34.1%; Pred. No. 8e-05;
Matches 30; Conservative 8; Mismatches 39; Indels 11; Gaps 4;

QY 31 AAHLEGGKDNQFIWKPQGWGRTGDCGGVQSVRAVWCFHVDG---WTSHLSNCGESNRP 87
DB 785 AIHPGCDKNCFAHWLAQDWERCNTTCG-RGVKKLVLCMELANGKQIRSGPEGGLAKRP 843
QY 88 PKERSCF-RVCDWHSDDLQWEVSDWHHC 114
DB 844 PEESTCFERPC-----FKWYTPWSEC 865

RESULT 10
O86TH1 PRELIMINARY; PRT; 951 AA.
AC O86TH1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA0605.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Peripheral Nervous System;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050544; AAH50544.1; -.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00090; tsp_1; 5.
DR SMART; SM00209; TSPI; 7.
DR PROSITE; PS0092; TSPI; 4.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 104620 MW; 93A9B0DC58AB6CC7 CRC64;

Query Match 14.4%; Score 119.5; DB 4; Length 951;
Best Local Similarity 34.1%; Pred. No. 8e-05;
Matches 30; Conservative 8; Mismatches 39; Indels 11; Gaps 4;

QY 31 AAHLEGGKDNQFIWKPQGWGRTGDCGGVQSVRAVWCFHVDG---WTSHLSNCGESNRP 87
DB 785 AIHPGCDKNCFAHWLAQDWERCNTTCG-RGVKKLVLCMELANGKQIRSGPEGGLAKRP 843
QY 88 PKERSCF-RVCDWHSDDLQWEVSDWHHC 114
DB 844 PEESTCFERPC-----FKWYTPWSEC 865
```



```

RESULT 11
ID O15916 PRELIMINARY; PRT; 724 AA.
AC O15916;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Random slug cDNA22 protein.
GN RSC22.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliales; Dictyostelium.
OC NCBI_TaxID=44683;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Iranfar N., Loomis W.F.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U82516; AAB69636.1; -.
DR InterPro; IPR009007; Rept_A_acid.
DR InterPro; IPR000884; TSP1.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS50092; TSP1; 2.
SQ SEQUENCE 724 AA; 81684 MW; 0D9B2528F6B6A8B6 CRC64;

Query Match 14.2%; Score 118; DB 5; Length 724;
Best Local Similarity 34.7%; Pred. No. 8.9e-05;
Matches 25; Conservative 9; Mismatches 34; Indels 4; Gaps 4;

QY 44 WK-PGPGWRCGCGGQVQSRVWCFHVDGWTSHLSNCGESNRPKPKSCFRCVCDWHS 102
DB 512 WKVGNWNSDCQICG-SGLQTRVLCTISINGELLDQERCNSHTKPPSERVC-AVAKCQAE 569

QY 103 LFOVEVSDWHHC 114
DB 570 -YHWD:SDWSG 580

RESULT 12
QYUF24 PRELIMINARY; PRT; 898 AA.
ID QYUF24
AC QYUF24;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP434H204.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Wambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL110226; CAB53684.1; -.
DR PIR; T14764; T14764.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 6.
DR SMART; SM00209; TSP1; 7.
DR PROSITE; PS50092; TSP1; 6.
DR KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 898 AA; 96731 MW; 38532F1E56F0A11B CRC64;

Query Match 14.2%; Score 117.5; DB 4; Length 898;
Best Local Similarity 29.8%; Pred. No. 0.00013;
Matches 33; Conservative 4; Mismatches 42; Indels 27; Gaps 4;

QY 44 WKPGWRCGTCDCGPGVQSRVWCFHVDGWTSHLS-----NCGESNRPKPKSCFR--VC 97
DB 95 WWAGEWLCSSCGPGGLSRAVLCTRSGLDQSALEPPACEHLPRPTTETPCNRHVC 154

```

```

QY 98 DWHSDLFQWEVSDWHHCVLVFPYARGVKPRTAECVTAQHGLQHR 141
DB 155 P-----ATWAVGNWSQC-----SVTCGEGTQRR 177

RESULT 13
QY1Z56 PRELIMINARY; PRT; 340 AA.
ID QY1Z56
AC QY1Z56;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to a disintegrin and metalloproteinase with thrombospondin
DE motifs 1 (ADAMTS-1) (Fragment).
GN ADAMTS15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009667; AAH09667.1; -.
DR MGD; MGI:2449669; Adams15.
DR CO; GO:0007223; P:integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 2.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS50092; TSP1; 2.
DR KW Integrin.
FT NON TER
SQ SEQUENCE 340 AA; 37021 MW; 06562C747634A8BD CRC64;

Query Match 14.0%; Score 116.5; DB 11; Length 340;
Best Local Similarity 32.8%; Pred. No. 5.9e-05;
Matches 40; Conservative 16; Mismatches 51; Indels 15; Gaps 8;

QY 18 RKLFLLSLLSHAHLEGGKDNQ--FIWKPGRCTGCGPGVQSRVWCFHVDGWT 75
DB 206 RGSVLNRNSVLNSNQVE-QPNRPPARWAGSWGPCVSCG-SGLQKRAVDCRDSFG-Q 262

QY 76 SHLSNCGESNRPKPKSCFRCVCDWHSDFQWEVSDWHHCVLVFPYARGVKPRTAECVTAQ 135
DB 263 QGASACDVDFHEPLEKACGEPCP-----TWELGNWSPC-SKSCRG-FKERPLKCV--G 312

QY 136 HG 137
DB 313 HG 314

RESULT 14
QYHBS6 PRELIMINARY; PRT; 237 AA.
ID QYHBS6
AC QYHBS6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel Human cDNA clones with function of inhibiting cancer cell
RT growth."
RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF217974; AAG17217.1; -.
DR Genew; HGNC:19706; TSR1.

```

```
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; TSP1; 2.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS00092; TSP1; 3.
KW Hypothetical protein.
SQ SEQUENCE 237 AA; 25659 MW; F873ED571107026C CRC64;

Query Match          14.0%; Score 116; DB 4; Length 237;
Best Local Similarity 26.6%; Pred. No. 4.6e-05;
Matches 29; Conservative 16; Mismatches 54; Indels 10; Gaps 4;

QY 36 GKKNQFTWKPGWGRCTGDCGGVQSRVAVWCFHVDGWTSHL---SNCGESNRPPPKERS 92
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
69 GPCERTWRWYTGPGWGECSSECG-SGTQRRDIIQVSKLGTETFTVSPNCSHLPRPP----123
QY 93 CFRVCDWHSDLFQWVSDWHCHLVLPYARGEVKPRTAECVTAQHGLQHR 141
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 ALQPCQGAQCDRWFTFWSPC--SRSCGGGTQTRVQCCLSTNQTLSR 170

RESULT 15
QSN643
ID Q8N643 PRELIMINARY; PRT; 424 AA.
AC Q8N643;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DI 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC027478; AAH27478.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; TSP1; 4.
DR SMART; SM00209; TSP1; 6.
DR PROSITE; PS00092; TSP1; 5.
SQ SEQUENCE 424 AA; 45774 MW; C4C51297C98C75EB CRC64;

Query Match          14.0%; Score 116; DB 4; Length 424;
Best Local Similarity 26.6%; Pred. No. 8.5e-05;
Matches 29; Conservative 16; Mismatches 54; Indels 10; Gaps 4;

QY 36 GKKNQFTWKPGWGRCTGDCGGVQSRVAVWCFHVDGWTSHL---SNCGESNRPPPKERS 92
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
256 GPCERTWRWYTGPGWGECSSECG-SGTQRRDIIQVSKLGTETFTVSPNCSHLPRPP----310
QY 93 CFRVCDWHSDLFQWVSDWHCHLVLPYARGEVKPRTAECVTAQHGLQHR 141
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
311 ALQPCQGAQCDRWFTFWSPC--SRSCGGGTQTRVQCCLSTNQTLSR 357
```

Search completed: February 24, 2004, 01:07:27  
Job time : 12.692 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

run on: February 24, 2004, 01:00:01 ; Search time 5.21344 Seconds  
(without alignments)  
1406.151 Million cell updates/sec

title: US-10-022-710-6\_COPY\_1\_142

effect score: 830

sequence: 1 MFPKSNLTVTCWVRMSRKL.....EVKPRTAECVTAQHGLQHRM 142

scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/prodata/2/1aa/5A-COMB.pep.\*
- 2: /cgn2\_6/prodata/2/1aa/5B-COMB.pep.\*
- 3: /cgn2\_6/prodata/2/1aa/6A-COMB.pep.\*
- 4: /cgn2\_6/prodata/2/1aa/6B-COMB.pep.\*
- 5: /cgn2\_6/prodata/2/1aa/PCITUS-COMB.pep.\*
- 6: /cgn2\_6/prodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	122	14.7	1882	4	US-09-369-364A-13
2	116.5	14.0	1745	4	US-09-800-729-89
3	114.5	13.8	1059	4	US-09-800-729-217
4	114	13.7	551	4	US-09-130-491-16
5	114	13.7	727	4	US-09-445-023A-12
6	109.5	13.2	1211	4	US-09-491-522-5
7	109.5	13.2	2165	4	US-09-800-729-155
8	106.5	12.8	1205	4	US-09-491-522-11
9	103	12.4	525	4	US-09-369-364A-21
10	100.5	12.1	1224	4	US-09-930-872-4
11	97.5	11.7	262	4	US-09-800-729-125
12	94	11.3	514	4	US-09-800-729-124
13	92	11.1	727	4	US-09-445-023A-1
14	92	11.1	949	4	US-09-568-559-2
15	92	11.1	967	4	US-09-130-491-2
16	88.5	10.7	450	4	US-09-369-364A-19
17	88.5	10.7	1081	4	US-09-369-364A-17
18	88	10.6	1184	2	US-08-918-914-1
19	88	10.6	1184	3	US-08-926-083-3
20	87	10.5	56	4	US-09-800-729-173
21	87	10.5	58	4	US-09-800-729-172
22	85.5	10.3	930	4	US-09-369-364A-2
23	85	10.2	142	4	US-09-380-882-8
24	84.5	10.2	59	4	US-09-800-729-163
25	83	10.0	481	4	US-09-130-491-8
26	83	10.0	481	4	US-09-352-991A-16955
27	81.5	9.8	518	4	US-09-369-364A-22

28 81.5 9.8 930 4 US-09-122-126B-15 Sequence 15, Appl  
29 81.5 9.8 930 4 US-09-634-286A-15 Sequence 15, Appl  
30 80 9.6 859 4 US-09-369-364A-5 Sequence 5, Appl  
31 79.5 9.6 788 2 US-08-918-914-4 Sequence 4, Appl  
32 78.5 9.5 608 4 US-09-130-491-13 Sequence 13, Appl  
33 78 9.4 59 4 US-09-800-729-171 Sequence 171, Appl  
34 77 9.3 757 4 US-09-963-791-24 Sequence 24, Appl  
35 77 9.3 905 4 US-09-369-364A-9 Sequence 9, Appl  
36 77 9.3 908 4 US-09-963-791-2 Sequence 2, Appl  
37 76.5 9.2 143 3 US-09-012-084-2 Sequence 2, Appl  
38 76.5 9.2 143 4 US-09-380-882-6 Sequence 6, Appl  
39 76.5 9.2 144 1 US-07-974-848B-2 Sequence 2, Appl  
40 76.5 9.2 837 4 US-09-122-126B-2 Sequence 2, Appl  
41 76.5 9.2 837 4 US-09-634-286A-2 Sequence 21, Appl  
42 74 8.9 144 4 US-09-180-100-21 Sequence 21, Appl  
43 74 8.9 153 4 US-09-252-991A-26199 Sequence 26199, A  
44 74 8.9 159 4 US-09-180-100-23 Sequence 23, Appl  
45 74 8.9 376 4 US-09-180-100-22 Sequence 22, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-369-364A-13  
; Sequence 13, Application US/09369364A  
; Patent No. 6391610  
; GENERAL INFORMATION:  
; APPLICANT: Apte, Suneel  
; APPLICANT: Hurskainen, Tiina L.  
; APPLICANT: Hirohata, Satoshi  
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases  
; FILE REFERENCE: 26473/4007/10-30-00  
; CURRENT APPLICATION NUMBER: US/09/369,364A  
; CURRENT FILING DATE: 1999-08-06  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1882  
; TYPE: PRT  
; ORGANISM: Homo sapiens ADAMTS-9  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (468)  
; OTHER INFORMATION: Xaa = C  
; NAME/KEY: MOD\_RES  
; LOCATION: (521)  
; OTHER INFORMATION: Xaa = Y  
US-09-369-364A-13

Query Match 14.7%; Score 122; DB 4; Length 1882;  
Best Local Similarity 34.7%; Pred. No. 0.00015;  
Matches 25; Conservative 12; Mismatches 25; Indels 10; Gaps 4;  
QY 44 WKQPGWCRCTGDCPGQVQSRVAVCFHVDGWTSHLSCGSGNRPKRS-C-FRVCDWHS 102  
DB 1279 WRTGFWGACSTCA-GGQRRVVVQDENGTYA--NDCVERIKPEQRCESGFCP---- 1331

QY 103 LFOEVSDDWHC 114

DB 1332 --QWYGNWGE 1341

##### RESULT 2

US-09-800-729-89  
; Sequence 89, Application US/09800729  
; Patent No. 6605592  
; GENERAL INFORMATION:  
; APPLICANT: Ni et al.  
; TITLE OF INVENTION: 32 Human secreted proteins  
; FILE REFERENCE: P204421  
; CURRENT APPLICATION NUMBER: US/09/800,729  
; CURRENT FILING DATE: 2001-03-08

;; PRIOR APPLICATION NUMBER: PCT/US00/26013  
;; PRIOR FILING DATE: 2000-09-22  
;; PRIOR APPLICATION NUMBER: 60/155,709  
;; PRIOR FILING DATE: 1999-09-24  
;; NUMBER OF SEQ ID NOS: 217  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 89  
;; LENGTH: 1745  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-800-729-89

Query Match 14.0%; Score 116.5; DB 4; Length 1745;  
Best Local Similarity 34.7%; Pred. No. 0.00054;  
Matches 33; Conservative 7; Mismatches 46; Indels 9; Gaps 5;  
QY 44 WKPGWGRCTGDC-GPG-GVQSRVAVCFHVDGWTSHLSNCGESNRPPKERSCF-RVCDWH 100  
DB 1593 WAFSSWGQCGPCIGPLAVQHRQVFCQTRDGTLPSEQCSALPRPVSTQNCWSEACSVH 1652  
QY 101 SDFQWEVSDWHHCVLVYPYARGEVKPRTACVTAQ 135  
DB 1653 -----WRVSLWTLCTATCGNYG-FQSRVVECVHAR 1681

RESULT 3  
US-09-800-729-217  
;; Sequence 217, Application US/09800729  
;; Patent No. 6605592  
;; GENERAL INFORMATION:  
;; APPLICANT: Ni et al.  
;; TITLE OF INVENTION: 32 Human secreted proteins  
;; FILE REFERENCE: P2044P1  
;; CURRENT APPLICATION NUMBER: US/09/800,729  
;; CURRENT FILING DATE: 2001-03-08  
;; PRIOR APPLICATION NUMBER: PCT/US00/26013  
;; PRIOR FILING DATE: 2000-09-22  
;; PRIOR APPLICATION NUMBER: 60/155,709  
;; PRIOR FILING DATE: 1999-09-24  
;; NUMBER OF SEQ ID NOS: 217  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 217  
;; LENGTH: 1059  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-800-729-217

Query Match 13.8%; Score 114.5; DB 4; Length 1059;  
Best Local Similarity 28.6%; Pred. No. 0.0005;  
Matches 34; Conservative 10; Mismatches 32; Indels 43; Gaps 8;  
QY 44 WKPGWGRCTGDCGPGVQSRVAVCFHVDGWTSHLSNCGESNRPPKERSCF----- 94  
DB 296 WEAGKWSKCTASCGQ-GVRRHVAC--VGG-----SDCDEGRPRQETTCYAGIPCSIAT 347  
QY 95 RVCWM-----HSDLFQWE-----VSDWHHCVLVYPYARGEVKPRTAECV 132  
DB 348 NSLDWMDRAYLDONTFGSMNDHND---WOAPRLVAGEWSTCSST---CGTGWMSRTVECV 401

RESULT 4  
US-09-130-491-16  
;; Sequence 15, Application US/09130491  
;; Patent No. 6416974  
;; GENERAL INFORMATION:  
;; APPLICANT: Holtzman, Douglas A.  
;; APPLICANT: Goodearl, Andrew D.J.  
;; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83  
;; FILE REFERENCE: 09404/041001  
;; CURRENT APPLICATION NUMBER: US/09/130,491  
;; CURRENT FILING DATE: 1998-08-07  
;; EARLIER APPLICATION NUMBER: US 60/058,108  
;; EARLIER FILING DATE: 1997-09-05

;; EARLIER APPLICATION NUMBER: US 60/054,961  
;; EARLIER FILING DATE: 1997-08-06  
;; NUMBER OF SEQ ID NOS: 16  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 16  
;; LENGTH: 551  
;; TYPE: PRT  
;; ORGANISM: Rattus rattus  
US-09-130-491-16

Query Match 13.7%; Score 114; DB 4; Length 551;  
Best Local Similarity 28.7%; Pred. No. 0.00027;  
Matches 37; Conservative 15; Mismatches 49; Indels 28; Gaps 8;  
QY 26 LLLSHA-----AHLEGGKDNQF-----TWKPGWGRCTGDCGPGVQSRVAVCFHV 71  
DB 410 LMVGHALRPKIKFYFMCKKTESFNAIPTFSEWIEEWGECSTCG-SGWQRRVVQCRDI 468  
QY 72 DGWTSHLNCGESNRPPKERSCFRVCDWHSDL--FQWEVSDWHHCVLVYPYARGEVKPRTA 129  
DB 469 NGHPA--SECAKEVKPASTRPC-----ADLPCPHWQVGDWSPCKT-CGKG-YKKRTL 517  
QY 130 ECVTAQHGL 138  
DB 518 KCVSHDGGV 526

RESULT 5  
US-09-445-023A-12  
;; Sequence 12, Application US/09445023A  
;; Patent No. 6565858  
;; GENERAL INFORMATION:  
;; APPLICANT: Hirose, Kunitaka  
;; APPLICANT: Inoguchi, Ei-ji  
;; APPLICANT: Hakoizaki, Michinozi  
;; APPLICANT: Ishicka, Keiko  
;; APPLICANT: Ishida, Yukako  
;; APPLICANT: Matsushima, Kouji  
;; APPLICANT: Kuno, Kouji  
;; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical  
;; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS  
;; FILE REFERENCE: Q57092  
;; CURRENT APPLICATION NUMBER: US/09/445,023A  
;; CURRENT FILING DATE: 1999-12-03  
;; PRIOR APPLICATION NUMBER: JP 9-160422  
;; PRIOR FILING DATE: 1997-06-03  
;; NUMBER OF SEQ ID NOS: 14  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 12  
;; LENGTH: 727  
;; TYPE: PRT  
;; ORGANISM: Mus sp.  
US-09-445-023A-12

Query Match 13.7%; Score 114; DB 4; Length 727;  
Best Local Similarity 28.7%; Pred. No. 0.00037;  
Matches 37; Conservative 15; Mismatches 49; Indels 28; Gaps 8;  
QY 26 LLLSHA-----AHLEGGKDNQF-----TWKPGWGRCTGDCGPGVQSRVAVCFHV 71  
DB 586 LMVGHALRPKIKFYFMCKKTESFNAIPTFSEWIEEWGECSTCG-SGWQRRVVQCRDI 644  
QY 72 DGWTSHLNCGESNRPPKERSCFRVCDWHSDL--FQWEVSDWHHCVLVYPYARGEVKPRTA 129  
DB 645 NGHPA--SECAKEVKPASTRPC-----ADLPCPHWQVGDWSPCKT-CGKG-YKKRTL 693  
QY 130 ECVTAQHGL 138  
DB 694 KCVSHDGGV 702

RESULT 6  
US-09-491-522-5

Sequence 5, Application US/09491522

Patent No. 6428998

GENERAL INFORMATION:

APPLICANT: Collige, Alain

APPLICANT: Lapiere, Charles M.

APPLICANT: Prockop, Darwin J.

TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,

TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds, LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2811

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/491,522

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/886,333

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Abrams, Samuel B.

REGISTRATION NUMBER: 30,605

REFERENCE/DOCKET NUMBER: 8389-0060-999

TELEPHONE: 650-493-4935

TELEFAX: 650-493-5556

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1211 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

JS-09-491-522-5

Query Match

Best Local Similarity 13.2%; Score 109.5; DB 4; Length 1211;

Matches 30; Conservative 12; Mismatches 38; Indels 27; Gaps 6;

40 NQFIWKPGWGRCTGCGGQVQSRVAVCFHV--DGWTS--HLSNCGESNRPPKERSCFR 95

914 SQPVVWTGWEPCSTCGRTGMQVRSVRCIQPLHDNTTSVHAKHCNDA-RPESRRACSR 972

96 -VCDWHSDLFPQWEVSDWHHCVLVPYARGEVKPRTASCVTAGHQLQHR 141

973 ELCPG-----RWRAGPWSQC-----SVTCGNGTQPR 998

RESULT 7

JS-09-800-729-155

Sequence 155, Application US/09800729

Patent No. 6605592

GENERAL INFORMATION:

APPLICANT: Ni et al.

TITLE OF INVENTION: 32 Human secreted proteins

FILE REFERENCE: P2044P1

CURRENT APPLICATION NUMBER: US/09/800,729

CURRENT FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: PCT/US00/26013

PRIOR FILING DATE: 2000-09-22

PRIOR APPLICATION NUMBER: 60/155,709

PRIOR FILING DATE: 1999-09-24

NUMBER OF SEQ ID NOS: 217

SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 155

; LENGTH: 2165

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-800-729-155

Query Match

Best Local Similarity 13.2%; Score 109.5; DB 4; Length 2165;

Matches 28; Conservative 13; Mismatches 32; Indels 25; Gaps 4;

QY 1 MFPSNLTVTCWVRSWRKLFLLLSLHAAHLEGGKDNQFIWKPGWGRCTGDCGPGG 60

Db 1865 LFPAPNESQTCEL-----NPCDSEFKWSTGFWGECNKCQ-G 1901

QY 61 VQSRVAVCFHVDTGWTSHLSNCGESNRPPKERSCF-RVC 97

Db 1902 IRRRRVKCVANDGRRVERVKC-TTKKPRRTQYCFERNC 1938

RESULT 8

US-09-491-522-11

Sequence 11, Application US/09491522

Patent No. 6428998

GENERAL INFORMATION:

APPLICANT: Collige, Alain

APPLICANT: Lapiere, Charles M.

APPLICANT: Prockop, Darwin J.

TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,

TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds, LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2811

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/491,522

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/886,333

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Abrams, Samuel B.

REGISTRATION NUMBER: 30,605

REFERENCE/DOCKET NUMBER: 8389-0060-999

TELEPHONE: 650-493-4935

TELEFAX: 650-493-5556

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 1205 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-09-491-522-11

Query Match

Best Local Similarity 12.8%; Score 106.5; DB 4; Length 1205;

Matches 30; Conservative 13; Mismatches 37; Indels 27; Gaps 6;

QY 40 NQFIWKPGWGRCTGDCGPGVQSRVAVVC---FHVDTGWTSHLSNCGESNRPPKERSCFR 95

Db 908 SQPVVWTGWEVFCRSRGRTGMQVRSVRCVQLHNNTTTSVHTKHCNDA-RPEGRACNR 966

```
QY 96 -VCDWHSDFQWVSDHHCVLPVYARGVKPRTAECVTAQHGLOHR 141
Db 967 ELCPG-----RWAGSQSC-----SVTCNGTQER 992

RESULT 9
US-09-369-364A-21
; Sequence 21, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Aptec, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirahata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-R1
US-09-369-364A-21

Query Match 12.4%; Score 103; DB 4; Length 525;
Best Local Similarity 25.2%; Pred. No. 0.004;
Matches 28; Conservative 14; Mismatches 27; Indels 42; Gaps 6;

QY 44 WKPGWGRCTGCGPGVQSRVAVWCFF-----VDGWTSHLNCGESNRPPKERSCF 94
Db 380 WEATPWTACSSCG-GGIQSRVAVSCVEEDIQHVTSVSEW-----KMYTPKMPIAQPC- 432

QY 95 RYCDWHSDF-----QWVSDHHCVLPVYARGVKPRTAECVTAQHGLOHR 142
Db 433 -----NIFDCPKLAQEWSPCT-----VTCGGLRYRV 460

RESULT 10
US-09-930-872-4
; Sequence 4, Application US/09930872
; Patent No. 6448388
; GENERAL INFORMATION:
; APPLICANT: Fiddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. 6448388el Human Proteases and Polynucleotides Encoding the Sa
; FILE REFERENCE: LEX-0219-USA
; CURRENT APPLICATION NUMBER: US/09/930,872
; PRIOR FILING DATE: 2001-08-14
; PRIOR FILING DATE: 2001-08-14
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1224
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-930-872-4

Query Match 12.1%; Score 100.5; DB 4; Length 1224;
Best Local Similarity 28.0%; Pred. No. 0.02;
Matches 30; Conservative 6; Mismatches 28; Indels 43; Gaps 6;

QY 44 WKPGWGRCTGCGPGVQSRVAVWCFFVDGWTSHLNCGESNRPPKERSCFVCDWH--- 100
Db 1130 WPAFSPNSQCTASCG-GGVOTRSVQCL---AGGRPASGCLLHQKP-----SASLACNTHFCP 1181

QY 101 -----SDLPQWVSDHHCVLPVYARGVKPRTAECVTAQHGL 138
Db 1182 IAEKDAFCNDYFHW-----CYLVP-----QHGML 1205

RESULT 11
```

```
US-09-800-729-125
; Sequence 125, Application US/09800729
; Patent No. 6605592
; GENERAL INFORMATION:
; APPLICANT: Ni et al.,
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (254)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-800-729-125

Query Match 11.7%; Score 97.5; DB 4; Length 262;
Best Local Similarity 29.4%; Pred. No. 0.0073;
Matches 40; Conservative 16; Mismatches 55; Indels 25; Gaps 10;

QY 20 LFLLLSLLSHAHLEGGKDNQFIWKP-GPWGRCTGDCGPGGVQS--RAVMCFHYDWTGS 76
Db 14 LFLAFLLSSRTARSEEDRDG--LWDANGPWSECSRTCGGASYSLLRCLSSKSCGRNI 71

QY 77 HLSNCGESNRPPKERSCFV--CDWHSDF-----FQW-EVSD--WHHCVLPYARG--- 122
Db 72 RYRCSNVDCPP-EAGDFRAQCSAHNDYKHGQFYEMLPVSDNDPNPCSLKQAKGTTL 130

QY 123 --EVKPRTAEE---CVT 133
Db 131 VVELAPKVIDGTTCYT 146

RESULT 12
US-09-800-729-124
; Sequence 124, Application US/09800729
; Patent No. 6605592
; GENERAL INFORMATION:
; APPLICANT: Ni et al.,
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 124
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-124

Query Match 11.3%; Score 94; DB 4; Length 514;
Best Local Similarity 33.3%; Pred. No. 0.038;
Matches 25; Conservative 6; Mismatches 34; Indels 10; Gaps 4;

QY 44 WKPGWGRCTGDCGPGVQSRVAVWCFF--VDGWTSHLNCGESNRPPKERSCFVCDW 99
Db 23 WYPAQWQPCSTCG-GGVQKREVLCKQRVADGSFLELPETFCAS-----KPACQQAACK 76

QY 100 HSDLFQWVSDHHC 114
```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

MM protein - protein search, using sw model

run on: February 24, 2004, 01:07:37 ; Search time 59.0268 Seconds  
(without alignments)  
507.969 Million cell updates/sec

Title: US-10-022-710-6\_COPY\_1\_142

Perfect score: 830

Sequence: 1 MFPPKGLNLTVCWVRMRKL.....EVKPTAEVTAQHGLQHRM 142

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10C\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	573	69.0	247	9	US-09-864-761-43221
2	327.5	39.5	933	15	US-10-104-047-2928
3	289	34.8	277	9	US-09-864-761-46564
4	252.5	30.4	1588	13	US-10-000-512-2
5	252.5	30.4	1588	15	US-10-074-566-2
6	252.5	30.4	1588	15	US-10-074-566-41
7	128.5	15.5	1057	14	US-10-188-869-10
8	128.5	15.5	1122	14	US-10-188-869-13
9	128.5	15.5	1145	14	US-10-188-869-20
10	127	15.3	1629	9	US-09-972-467-2
11	127	15.3	1907	9	US-09-938-330-25
12	126	15.2	1916	15	US-10-274-639-10
13	126	15.2	1280	14	US-10-087-887-86
14	122	14.7	1882	9	US-09-918-171A-13
15	121.5	14.6	226	9	US-09-945-676-8

Sequence 12, Appli  
Sequence 6, Appli  
Sequence 10, Appli  
Sequence 4, Appli  
Sequence 8, Appli  
Sequence 2, Appli  
Sequence 124, Appl  
Sequence 6, Appli  
Sequence 47, Appli  
Sequence 2, Appli  
Sequence 5, Appli  
Sequence 762, Appl  
Sequence 89, Appli  
Sequence 222, Ap  
Sequence 2, Appli  
Sequence 554, Appl  
Sequence 45, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 10, Appli  
Sequence 217, Appl  
Sequence 10, Appli  
Sequence 16, Appli  
Sequence 16, Appli  
Sequence 12, Appli  
Sequence 12, Appli  
Sequence 4, Appli  
Sequence 7, Appli  
Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-864-761-43221  
; Sequence 43221, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30



48 PWGRCTGD-CGPGGVQSRVAVWCFFHYDGTWTSLSNCGESNRPPKERSCFRVCWHDWSDLFQW 106

```
Db 1 PWGRCMDECGPGGIQTRAVWCARHVEGWTTLTNCQAERPNNQCNCFKVCDDWHKELYDM 60
2Y 107 EVSDWHCHVLVPYARGEVKPRTAECVTAGHLOHR 141
Db 61 RLGPWQCQPV-ISKSLKP--LECIKGEGIQVR 92

RESULT 4
US-10-000-512-2
; Sequence 2, Application US/10000512
; Publication No. US20020164699A1
; GENERAL INFORMATION:
; APPLICANT: Shinkets, Richard A
; APPLICANT: Fernandes, Elma
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME
; FILE REFERENCE: 15966-556
; CURRENT APPLICATION NUMBER: US/10/000,512
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/619,252
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/167,785
; PRIOR FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1588
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-000-512-2

Query Match 30.4%; Score 252.5; DB 13; Length 1588;
Best Local Similarity 44.8%; Pred. No. 3.7e-17;
Matches 39; Conservative 21; Mismatches 24; Indels 3; Gaps 2;

2Y 55 DCGPGGVQSRVAVWCFFHVDGWTSHLSNCGESNRPKPSRCFVCDWHSDDLFWQEVSDWHHC 114
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 4 ECGPGGIQTRAVWCARHVEGWTTLTNCQAERPNNQCNCFKVCDDWHKELYDMRLGPWQC 63
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2Y 115 VLPVYARGEVKPRTAECVTAGHLOHR 141
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 64 QPV-ISKSLKP--LECIKGEGIQVR 87
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 5
US-10-074-566-2
; Sequence 2, Application US/10074566
; Publication No. US20030207348A1
; GENERAL INFORMATION:
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Shenoy, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same
; FILE REFERENCE: 15966-556 CIP1
; CURRENT APPLICATION NUMBER: US/10/074,566
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: 09/619,252
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/144,722
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/167,785
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 60/276,994
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/280,898
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/332,241
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/288,062
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/291,766
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/314,007
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 1588
; TYPE: PRT
; ORGANISM: human
US-10-074-566-41

Query Match 30.4%; Score 252.5; DB 15; Length 1588;
Best Local Similarity 44.8%; Pred. No. 3.7e-17;
```

```
; PRIOR APPLICATION NUMBER: 60/288,062
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/291,766
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/314,007
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1588
; TYPE: PRT
; ORGANISM: human
US-10-074-566-2

Query Match 30.4%; Score 252.5; DB 15; Length 1588;
Best Local Similarity 44.8%; Pred. No. 3.7e-17;
Matches 39; Conservative 21; Mismatches 24; Indels 3; Gaps 2;

QY 55 DCGPGGVQSRVAVWCFFHVDGWTSHLSNCGESNRPKPSRCFVCDWHSDDLFWQEVSDWHHC 114
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 4 ECGPGGIQTRAVWCARHVEGWTTLTNCQAERPNNQCNCFKVCDDWHKELYDMRLGPWQC 63
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 115 VLPVYARGEVKPRTAECVTAGHLOHR 141
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 64 QPV-ISKSLKP--LECIKGEGIQVR 87
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 6
US-10-074-566-41
; Sequence 41, Application US/10074566
; Publication No. US20030207348A1
; GENERAL INFORMATION:
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Fernandes, Elma R.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Spytek, Kimberly A.
; TITLE OF INVENTION: Polypeptides and Polynucleotides Encoding Same
; FILE REFERENCE: 15966-556 CIP1
; CURRENT APPLICATION NUMBER: US/10/074,566
; CURRENT FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: 09/619,252
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/144,722
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/167,785
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 60/276,994
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/280,898
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/332,241
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/288,062
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/291,766
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/314,007
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 1588
; TYPE: PRT
; ORGANISM: human
US-10-074-566-41

Query Match 30.4%; Score 252.5; DB 15; Length 1588;
Best Local Similarity 44.8%; Pred. No. 3.7e-17;
```

Matches 39; Conservative 21; Mismatches 24; Indels 3; Gaps 2;  
QY 55 DCGPGVQSVRAVCFHVDGWTSHLNCGESNRPPKRSCTFRVCDWHSDFQWEVSDWHHC 114  
Db 4 ECGPGIGITRAVCAHVEGWTILHNCQAEPNNQNCQCFKVCWDHKLIDYRLGPNQC 63  
QY 115 VLVFVARGVKPRTAEVTAQHLQHR 141  
Db 64 QPV-ISKSLKP--LECIKGEEGIQVR 87

RESULT 7  
US-10-188-869-10  
; Sequence 10, Application US/10188869  
; Publication No. US20030148306A1  
; GENERAL INFORMATION:  
; APPLICANT: LAVALLIE, EDWARD  
; APPLICANT: RACIE, LISA  
; APPLICANT: DIBLASIO, ELIZABETH  
; APPLICANT: AGOSTINO, MICHAEL  
; TITLE OF INVENTION: AGGREGANASE MOLECULES  
; FILE REFERENCE: 08702.0092-00000  
; CURRENT APPLICATION NUMBER: US/10/188,869  
; PRIOR FILING DATE: 2002-07-05  
; PRIOR APPLICATION NUMBER: 60/349,133  
; PRIOR FILING DATE: 2002-01-16  
; PRIOR APPLICATION NUMBER: 60/303,051  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 1057  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-188-869-10

Query Match 15.5%; Score 128.5; DB 14; Length 1057;  
Best Local Similarity 31.1%; Pred. No. 0.00014;  
Matches 33; Conservative 16; Mismatches 40; Indels 17; Gaps 7;  
QY 35 EGKDNQFTWKGPGWRCCTGDCGPGVQSVRAVCFHVDGWTSHLNCGESNRPPKRSCTF 94  
Db 879 EG-QDCLSIWEASEWSQCSASCGK-GWVKRTVAC-----TNSQKCDASTRPAEEA-- 928  
QY 95 RVCDSHSDLFQWEVSDWHHCVLVPYARGVKPRTAEC---VTAQHG 137  
Db 929 --CEDYSGCYEWKTDWSTCSST-CGKG-LQSRVVCQMKVGTGRHG 970

RESULT 8  
US-10-188-869-13  
; Sequence 13, Application US/10188869  
; Publication No. US20030148306A1  
; GENERAL INFORMATION:  
; APPLICANT: LAVALLIE, EDWARD  
; APPLICANT: RACIE, LISA  
; APPLICANT: DIBLASIO, ELIZABETH  
; APPLICANT: AGOSTINO, MICHAEL  
; TITLE OF INVENTION: AGGREGANASE MOLECULES  
; FILE REFERENCE: 08702.0092-00000  
; CURRENT APPLICATION NUMBER: US/10/188,869  
; CURRENT FILING DATE: 2002-07-05  
; PRIOR APPLICATION NUMBER: 60/349,133  
; PRIOR FILING DATE: 2002-01-16  
; PRIOR APPLICATION NUMBER: 60/303,051  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 1122  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-188-869-13

Query Match 15.5%; Score 128.5; DB 14; Length 1122;  
Best Local Similarity 31.1%; Pred. No. 0.00015;  
Matches 33; Conservative 16; Mismatches 40; Indels 17; Gaps 7;  
QY 35 EGKDNQFTWKGPGWRCCTGDCGPGVQSVRAVCFHVDGWTSHLNCGESNRPPKRSCTF 94  
Db 944 EG-QDCLSIWEASEWSQCSASCGK-GWVKRTVAC-----TNSQKCDASTRPAEEA-- 993  
QY 95 RVCDSHSDLFQWEVSDWHHCVLVPYARGVKPRTAEC---VTAQHG 137  
Db 994 --CEDYSGCYEWKTDWSTCSST-CGKG-LQSRVVCQMKVGTGRHG 1035

RESULT 9  
US-10-188-869-20  
; Sequence 20, Application US/10188869  
; Publication No. US20030148306A1  
; GENERAL INFORMATION:  
; APPLICANT: LAVALLIE, EDWARD  
; APPLICANT: RACIE, LISA  
; APPLICANT: DIBLASIO, ELIZABETH  
; APPLICANT: AGOSTINO, MICHAEL  
; TITLE OF INVENTION: AGGREGANASE MOLECULES  
; FILE REFERENCE: 08702.0092-00000  
; CURRENT APPLICATION NUMBER: US/10/188,869  
; CURRENT FILING DATE: 2002-07-05  
; PRIOR APPLICATION NUMBER: 60/349,133  
; PRIOR FILING DATE: 2002-01-16  
; PRIOR APPLICATION NUMBER: 60/303,051  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 1145  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-188-869-20

Query Match 15.5%; Score 128.5; DB 14; Length 1145;  
Best Local Similarity 31.1%; Pred. No. 0.00016;  
Matches 33; Conservative 16; Mismatches 40; Indels 17; Gaps 7;  
QY 35 EGKDNQFTWKGPGWRCCTGDCGPGVQSVRAVCFHVDGWTSHLNCGESNRPPKRSCTF 94  
Db 967 EG-QDCLSIWEASEWSQCSASCGK-GWVKRTVAC-----TNSQKCDASTRPAEEA-- 1016  
QY 95 RVCDSHSDLFQWEVSDWHHCVLVPYARGVKPRTAEC---VTAQHG 137  
Db 1017 --CEDYSGCYEWKTDWSTCSST-CGKG-LQSRVVCQMKVGTGRHG 1058

RESULT 10  
US-09-972-467-2  
; Sequence 2, Application US/09972467  
; Patent No. US20020090373A1  
; GENERAL INFORMATION:  
; APPLICANT: PFIZER INC.  
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND  
; FILE REFERENCE: PCL0850A  
; CURRENT APPLICATION NUMBER: US/09/972,467  
; CURRENT FILING DATE: 2001-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1629  
; TYPE: PRT  
; ORGANISM: Human  
US-09-972-467-2

Query Match 15.3%; Score 127; DB 9; Length 1629;  
Best Local Similarity 31.5%; Pred. No. 0.00032;

Matches 28; Conservative 15; Mismatches 34; Indels 12; Gaps 5;  
2Y 44 WKPGWRCCTGDCGPGVQSVRAYWCFHVDGWTSHLNCGSGNRPPKERSC-FRVCDWHS 102  
Db 1332 WRTGPGWACSTCA-GGSQRRVVVCQDENGTYA--NDCVERIKPDQRACSGPCP---- 1384  
2Y 103 LFWQVSDWHHCVLVVPYARGVPRTAEC 131  
Db 1385 --QWAYGNWGECTKL--CGGGIRTLVVC 1409  
RESULT 11  
US-09-938-330-25  
; Sequence 25, Application US/09938330  
; Patent No. US20020115838A1  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Hilbun, Brin  
; APPLICANT: Scoville, John  
; APPLICANT: Friddle, Carl Johan  
; APPLICANT: Hu, Yi  
; APPLICANT: Turner, C. Alexander Jr.  
; TITLE OF INVENTION: No. US20020115838A1 Human Proteases and Polynucleotides Encodin  
; FILE REFERENCE: LEX-0237-USA  
; CURRENT APPLICATION NUMBER: US/09/938,330  
; PRIOR APPLICATION NUMBER: 2001-08-22  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR FILING DATE: 2000-09-19  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 1907  
; TYPE: PRT  
; ORGANISM: homo sapiens  
JS-09-938-330-25  
Query Match 15.3%; Score 127; DB 9; Length 1907;  
Best Local Similarity 31.5%; Pred. No. 0.00038;  
Matches 28; Conservative 15; Mismatches 34; Indels 12; Gaps 5;  
2Y 44 WKPGWRCCTGDCGPGVQSVRAYWCFHVDGWTSHLNCGSGNRPPKERSC-FRVCDWHS 102  
Db 1304 WRTGPGWACSTCA-GGSQRRVVVCQDENGTYA--NDCVERIKPDQRACSGPCP---- 1356  
2Y 103 LFWQVSDWHHCVLVVPYARGVPRTAEC 131  
Db 1357 --QWAYGNWGECTKL--CGGGIRTLVVC 1381  
RESULT 12  
US-10-274-639-10  
; Sequence 10, Application US/10274639  
; Publication No. US20030232349A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: DELEGANE, Angelo M.; GANDHI, Ameena R.  
; APPLICANT: HAPALIA, April J.A.; LU, Dyung Aina M.  
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.  
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.  
; APPLICANT: NGUYEN, Darnell B.; LEE, Ernestine A.  
; APPLICANT: KHAN, Farrah A.; YUE, Henry  
; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.  
; APPLICANT: POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi  
; APPLICANT: YANG, Junming; THANGAVELU, Kavitha  
; APPLICANT: DING, Li; KEARNEY, Liam  
; APPLICANT: BAUGHN, Marian R.; BOROWSKY, Mark L.  
; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.  
; APPLICANT: BURFORD, Neil; WALLIA, Narinder K.  
; APPLICANT: LAL, Preeti G.; LEE, Sally  
; APPLICANT: TODD, Stephen; LO, Terence P.  
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.

; APPLICANT: AZIMZAI, Yalda; LU, Yan  
; TITLE OF INVENTION: PROTEASES  
; FILE REFERENCE: PI-0167 USA  
; CURRENT APPLICATION NUMBER: US/10/274,639  
; CURRENT FILING DATE: 2002-10-18  
; PRIOR APPLICATION NUMBER: PCT/US01/22397  
; PRIOR FILING DATE: 2001-07-17  
; PRIOR APPLICATION NUMBER: US 60/220,063  
; PRIOR FILING DATE: 2000-07-21  
; PRIOR APPLICATION NUMBER: US 60/221,680  
; PRIOR FILING DATE: 2000-07-28  
; PRIOR APPLICATION NUMBER: US 60/223,544  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/224,717  
; PRIOR FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: US 60/225,988  
; PRIOR FILING DATE: 2000-08-16  
; PRIOR APPLICATION NUMBER: US 60/227,568  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: PERL Program  
; SEQ ID NO 10  
; LENGTH: 1916  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No. US20030232349A1 2994162CD1  
US-10-274-639-10  
Query Match 15.3%; Score 127; DB 15; Length 1916;  
Best Local Similarity 31.5%; Pred. No. 0.00038;  
Matches 28; Conservative 15; Mismatches 34; Indels 12; Gaps 5;  
QY 44 WKPGWRCCTGDCGPGVQSVRAYWCFHVDGWTSHLNCGSGNRPPKERSC-FRVCDWHS 102  
Db 1313 WRTGPGWACSTCA-GGSQRRVVVCQDENGTYA--NDCVERIKPDQRACSGPCP---- 1365  
QY 103 LFWQVSDWHHCVLVVPYARGVPRTAEC 131  
Db 1366 --QWAYGNWGECTKL--CGGGIRTLVVC 1390  
RESULT 13  
US-10-087-887-86  
; Sequence 86, Application US/10087887  
; Publication No. US20030198957A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Conley, Pamela B.  
; APPLICANT: Yang, Ruey-Bing  
; APPLICANT: Hart, Matthew  
; APPLICANT: Tomlinson, James E.  
; APPLICANT: Topper, James N.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Zethusen, Bryan D.  
; APPLICANT: Komuves, Laszlo  
; APPLICANT: Padigaru, Muralidhara  
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 21402-285  
; CURRENT APPLICATION NUMBER: US/10/087,887  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: 60/273,049  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: 60/279,883  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/277,791  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/281,248  
; PRIOR FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: 60/282,864  
; PRIOR FILING DATE: 2001-04-10

;; PRIOR APPLICATION NUMBER: 60/282,537  
;; PRIOR FILING DATE: 2001-04-09  
;; PRIOR APPLICATION NUMBER: 60/282,867  
;; PRIOR FILING DATE: 2001-04-10  
;; NUMBER OF SEQ ID NOS: 104  
;; SOFTWARE: CuraseqList version 0.1  
;; SEQ ID NO 86  
;; LENGTH: 1280  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-10-087-887-86

Query Match 15.2%; Score 126; DB 14; Length 1280;  
Best Local Similarity 32.6%; Pred. No. 0.00032;  
Matches 31; Conservative 15; Mismatches 39; Indels 10; Gaps 6;  
QY 42 FIWKPGWGRCTGDCGGVQSVRAVMC-FHVDGWTSHLSNCGESNRPPKERSC-FRVCDW 99  
Db 307 FSWSHSGWGDCAECG-GGHQSLVFTIDNEAYPDH--CQHQRPTHRRSCNTQPC-- 361  
QY 100 HSDLFQWEVSDWHHCVLVYPYARGVKPRTAECVTA 134  
Db 362 -PKTKRWKVPWTPCSV--SCGGVQSVRSVYCISS 393

## RESULT 14

US-09-918-171A-13  
;; Sequence 13, Application US/09918171A  
;; Patent No. US20020110894A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Apte, Suneel  
;; APPLICANT: Hurskainen, Tiina L.  
;; APPLICANT: Hirohata, Satoshi  
;; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases  
;; FILE REFERENCE: 26473/04193  
;; CURRENT APPLICATION NUMBER: US/09/918,171A  
;; CURRENT FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: 09/369,364  
;; PRIOR FILING DATE: 1999-08-06  
;; NUMBER OF SEQ ID NOS: 31  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 13  
;; LENGTH: 1882  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens ADAMTS-9  
;; FEATURE:  
;; NAME/KEY: MOD RES  
;; LOCATION: (468)  
;; OTHER INFORMATION: Xaa = Cys  
;; NAME/KEY: MOD RES  
;; LOCATION: (521)  
;; OTHER INFORMATION: Xaa = Tyr  
US-09-918-171A-13

Query Match 14.7%; Score 122; DB 9; Length 1882;  
Best Local Similarity 34.7%; Pred. No. 0.0012;  
Matches 25; Conservative 12; Mismatches 25; Indels 10; Gaps 4;  
QY 44 WKPGWGRCTGCGGVQSVRAVMCFHVDGWTSHLSNCGESNRPPKERSC-FRVCDWHS 102  
Db 1279 WRTGPGWACSSICA-GGSQRRVVVQDENGTYA--NDCVERIKPDEQACBSGCP----- 1331  
QY 103 LFQWEVSDWHHC 114  
Db 1332 --QWAYGNWGE 1341

## RESULT 15

US-09-945-676-8  
;; Sequence 8, Application US/09945676  
;; Patent No. US20020077465A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Shi et al.

;; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies  
;; FILE REFERENCE: PT051P1  
;; CURRENT APPLICATION NUMBER: US/09/945,676  
;; CURRENT FILING DATE: 2001-09-05  
;; PRIOR APPLICATION NUMBER: PCT/US01/05497  
;; PRIOR FILING DATE: 2001-02-22  
;; PRIOR APPLICATION NUMBER: 60/187,937  
;; PRIOR FILING DATE: 2000-03-03  
;; NUMBER OF SEQ ID NOS: 9  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 8  
;; LENGTH: 226  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-945-676-8

Query Match 14.6%; Score 121.5; DB 9; Length 226;  
Best Local Similarity 28.2%; Pred. No. 0.00015;  
Matches 3; Conservative 19; Mismatches 43; Indels 17; Gaps 7;  
QY 31 AAHLEGGKDNQFTWKPGWGRCTGDCGGVQSVRAVMCFHVDGWTSHLSNCGESNRPPKE 90  
Db 47 AQCEG-QDCMTWEAGWSECSVKCGK-GIRHRTVRC-----INPRKKCVLSTPREA 98  
QY 91 RSCFRVCDWHSDDLQWEVSDWHHCVLVYPYARGVKPRTAEC---VTAQHG 137  
Db 99 ED----CEDYSKCYVVRMGDWSKC-SITCGKG-MQSRVIOCMHKITGRHG 142

Search completed: February 24, 2004, 01:30:05  
Job time : 59.0268 secs